nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

KIDNEY IN A DISH

A human organoid with a full set of renal cell types

MODERN LIFE

WHAT'S THE RUSH?

Diary lab reveals why life is so hectic

POLLUTION

UNPICK CHINA'S CHOKING MIX

Noxious cocktail mystifies atmospheric chemists

ASTROPHYSICS

THE DYING OF THE LIGHT

A disintegrating planet orbiting a white dwarf

Russian roulette

Attempts to keep foreign interests out of Russian research will only suppress the exchange of information, and risk damaging East–West relations.

Despite decades of intellectual isolation, the Soviet Union produced some fine science. When it imploded, only a wave of foreign aid and philanthropy protected that excellent research base from collapse. The strategy worked: as individualism and entrepreneurship took hold in Russia, science regained its strength and started to look outwards — as any successful research endeavour must in the twenty-first century.

Yet Russian President Vladimir Putin believes that his country can increasingly go its own way, and centralism and anti-Western rhetoric are on the rise. Science is beginning to suffer from paranoid state control.

As we report on page 486, Russia has placed strict new rules on how its scientists can operate. In response to a recently amended law, Russian universities and research institutes have begun to instruct scientists to seek permission from the Federal Security Service before they submit papers or give talks at scientific conferences..

The wording of the law is vague, seemingly deliberately so. It effectively requires any work that is applicable to industry to be approved for publication. Russian scientists are rightly outraged by this return to inglorious Soviet practices.

Meanwhile, dozens of organizations that receive foreign funding (and which the Russian government suspects are involved in "political activities" — again vaguely defined) are under scrutiny. Officially, this is to identify and repel unwelcome foreign influence. Unofficially, there is a whiff of political scores being settled.

In May, the Dynasty Foundation, Russia's largest private science-funding organization, shut down after the Ministry of Justice labelled it a "foreign agent". Other philanthropic groups and foreign-funded foundations fear that they may soon find themselves on a list of "undesirable" organizations that the Russian parliament is drawing up.

This is not the 1960s. Today, fear and isolationism can only damage collaborative science. In turn, this will undermine Russia's efforts to modernize its struggling economy. Putin knows only too well that his country's dependence on oil and gas exports is a treacherous anachronism as the world steers away from fossil-fuel use. Wisely, the government has substantially stepped up its science funding in recent years. But neither a multibillion-rouble nanotechnology initiative, launched in 2007, nor attempts to create a number of world-class research universities and attract top Western scientists to Russian labs will bear fruit if fear and distrust continue to stand in the way of a liberal science culture.

Russia's annexation of the Crimean Peninsula last year, and its dubious role in the ongoing conflict in the rest of Ukraine, chilled East–West collaborations, in science and other fields. Russia's controversial military involvement in the civil war in Syria, although cautiously tolerated by Western powers, threatens to cause further tension.

Through large European research facilities such as the particle-physics laboratory CERN and the international nuclear-fusion project ITER, science can still offer a much-needed peaceful counterbalance in these politically turbulent times. But a disturbingly anti-Western speech to the upper chamber of the Russian parliament by Putin's top science adviser on 30 September — the same day that Russia began its air strikes in Syria — testifies to the level of misunderstanding that is currently poisoning East–West relations across the board.

The speech by Mikhail Kovalchuk, director of the Kurchatov Institute of nuclear science in Moscow and a key contact for many international collaborations, delivered a patently absurd account, riddled with lies and propaganda, of how international science is a US plot to undermine Russia. Such anti-Western sentiments are readily echoed in Russia: last week, a high-ranking IT adviser to the government said that Russia should stop training computer experts because they will before long be serving Western interests.

> *"A crackdown on academic freedom and foreign support will be devastating."*

Making a bogeyman of the outside world — and in particular of the United States — is a populist political strategy intended to prepare the ground for anti-liberal isolationism. For Russia's scientific community, a crackdown on academic freedom and foreign support will be devastating. Putin, who frequently expresses his appreciation of science, must see that investment alone is not enough.

To pour cash into a system that stifles intuition, brilliance and truth will not help a nation that has always held scientists and explorers in great esteem. Even through difficult economic and political times, Russian science has produced a never-ending supply of great minds. It needs the freedom and respect to continue to do so. ∎

Abstract thoughts

Scientists, meeting organizers and the media must take care with preliminary findings.

The rough and tumble of professional science is no place for the faint-hearted. Progress rests on honest appraisal of methods and results. Ideas must be challenged and conclusions defended. One of the most important transitions for any researcher is swapping the textbook scrutiny of the undergraduate years for critical and creative thinking. At the centre of this culture is the academic conference.

Often the first chance for studies to be presented, discussed and criticized, these meetings are an important testing ground for early research. The community gets a heads-up on what others are doing, and how, and the scientists involved get some robust feedback that can shape their work.

Against such criteria, the presentation of preliminary data from a search for the genetic roots of homosexuality, at a meeting of the American Society of Human Genetics in Baltimore, Maryland, earlier this month, was a success. So why does it feel as if something went wrong?

In a ten-minute talk at the meeting, lead researcher Tuck Ngun described how his team scanned the DNA of 37 pairs of identical twins for chemical, or epigenetic, tags. They found a handful of similarities between many of the gay twins that were not present in their straight brothers.

Epigenetic tags, which often regulate gene expression, can be both inherited and affected by environmental factors, as seems to be the case for homosexuality itself. The findings were preliminary, but the idea that epigenetics is involved in sexual orientation is certainly plausible, and the researchers hoped that their findings would stimulate future research. Most labs shy away from studying homosexuality because funders are reluctant to wade into the topic and because of the well-founded worry that findings will be used in the misguided search for a 'cure'.

A flurry of press coverage ensued. Although some of the stories noted the study's small sample size and need for replication — limitations that the researchers readily acknowledged — others were somewhat less than circumspect. 'Have They Found the Gay Gene? Breakthrough in the US', screamed the front page of one newspaper.

Responding to the press coverage, many commentators took aim at the science — or at least what science was available in the 368-word conference abstract. The statistical analyses that the authors used are controversial, and there is a legitimate debate to be had. But short on hard information, the criticism turned into attack.

A few critics went so far as to argue that the authors should not have presented such preliminary work at the meeting. And at least one suggested that the authors could have provided preprints of their study when presenting it. These arguments seem to misunderstand the traditional, and still useful and relevant, role of such gatherings. Studies with small sample sizes and controversial methods are presented at conferences all the time, and many scientists already fear being scooped when they present even a bit of their data.

"Meetings are an important testing ground for early research."

It is unlikely that most newspapers seek science stories by meticulously scanning the abstract lists for foreign scientific conferences. It is much more likely that the wide coverage afforded to the epigenetics study arose because the story was presented to news desks in a press release from the conference organizers — and this is where there are lessons to learn.

The press release, which was not seen or approved by all the scientists involved, was titled 'Epigenetic Algorithm Accurately Predicts Male Sexual Orientation'. It certainly added to the potential for the study to be misinterpreted. The organizers have pledged to reconsider how they select which conference talks to highlight before a meeting, and how press releases are approved.

The genetics of homosexuality is a subject that will always find media coverage, partly because of the societal interest in the topic. Neither the scientists nor the conference organizers can be held responsible for how some in the media chose to write about the study. But both could have done more to get the right message across. ■

Pick and mix

Food regulators are right to place new forms of data on the safety menu.

Italian chocolate, Bangladeshi samosas, Chilean cornbread flans, Turkmenistani beef chapattis — the aromas of the world's traditional foods mingle seductively along the mile of pavilions at Expo Milano 2015, this year's world fair, dedicated to food. All delicious, but are they all safe? Will future foods be safe? Who is to judge — and on what evidence?

In Europe, the European Food Safety Authority (EFSA) decides whether a new food can be marketed, and its job (like that of all similar regulatory agencies around the world) is getting tougher. Technological advances are creating ever more novel foods.

The same technologies, along with the Internet and databases, have created more sources of information that may have a bearing on safety assessment: terabytes of molecular information from genomic or proteomic analyses, for example, or more-qualitative data generated through crowdsourcing.

Public trust in EFSA's decisions is patchy and, until now, the agency has been slow to engage with the problems and solutions that these technologies offer. But at a three-day conference in Milan — attached to the Expo, and concluding on 16 October, World Food Day — EFSA announced a new commitment to take on the modern challenges. As it does so, it can start to repair its rather undeserved reputation for non-transparency.

Created in 2002 and based in Parma, Italy, the agency is probably best known as the independent scientific advisory agency to the European Union, whose independent scientific advice on the safety of genetically modified (GM) cereals has been serially rejected by many EU member states.

In most cases, EFSA's science-based recommendations on the safety of new food products are accepted politically without too many questions. But the GM saga has encouraged a public distrust in its official scientific expertise. The scientific experts commissioned by EFSA over the years to analyse data on whether GM technologies or products are risky to health or the environment have seen their recommendations challenged time and again by protest groups that claim to have new data on dangers. As a one-off exception to the single-market rule, EU member states can decide on an individual basis whether they want to allow cultivation of a particular crop. Nineteen have registered their decisions to opt out, despite EFSA's seal of safety.

EFSA does a good job of risk assessment and is reasonably transparent — but to stop distrust from seeping into all areas of its work it needs to do more. Risk assessment is a complicated science to convey to the public and is becoming even more complex with every new potential source of information. EFSA must be transparent about the exact data that it uses to make individual judgements and about the methods it uses to determine the degree of uncertainty around those judgements. It must also find ways to transparently assign appropriate weight to different data types that have been collected with varying degrees of scientific rigour.

The agency is on the case. This year, it carried out a public consultation on the communication of uncertainties, and it is rolling out a toolbox of methods to be systematically tested over the next year. Such methods may address, for example, how to weigh up evidence generated from computer modelling, from animal data generated in labs or from data gathered over social media — or how to assess whether a particular change observed in an organism is biologically relevant.

By definition, risk assessment will never be able to deliver simple answers. And concerned citizens, rightly, will never place blind trust in scientific expertise. That is why transparency about both data sources and analysis methods is so important. Different people may even interpret the same complex data set differently. Citizens just need to be given a clear picture of how a risk assessor has interpreted data — so that they can challenge or accept the final decision of the risk manager. ■

↻ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

Indigenous peoples must benefit from science

To drive sustainable development, **Dyna Rochmyaningsih** *argues, science must empower rural communities — not just serve industry and governments.*

The sun has been pale for months here in Sumatra and the skies are grey all day — choked with pollution from the massive fires that rage across the Indonesian island. Since the late 1990s, the haze caused by these annual fires has posed a significant threat to the health of Sumatra's rural communities. This year's haze is especially bad and has affected major cities, both here and abroad; consequently, the fires have again made headlines around the world.

Many of these news stories blame the big palm-oil companies for the fires. Slash-and-burn techniques remain the cheapest way to clear forest for new plantations. But scientific evidence suggests that this simple narrative is not absolutely true. A number of surveys have found that the bulk of these fires are started outside the official oil-palm concessions. Small-scale farmers seem to be more to blame.

The haze in Indonesia is not just an environmental issue; it is a complex socio-economic problem that is driven partly by conflict over land ownership between palm-oil companies and rural communities — a struggle that the companies usually win.

Besides holding financial and legal power, these companies also have science on their side. High-quality research at state-funded centres has found ways to increase the production of palm oil, such as the manipulation of the gene *SHELL* and ways to weed out oil-palm clones with reduced yields. These technologies have been developed by the Malaysian Palm Oil Board, and the big companies in the region can pay to license and use them. But such technologies are out of the reach of small-holders and the rural population. Yet smallholders produce a large proportion of the crops, mainly through conventional farming practices.

Some 80% of Indonesian rubber, for example, is made by small-scale farmers who do not have access to the research products and whose welfare has not improved. What has science done to empower these people?

The problems of Indonesian farmers might seem low on the list of global priorities. But as the nations of the world prepare to discuss a treaty on climate change in Paris next month, the fires that fuel the Sumatran haze offer a perfect example of how the relationship between science and industry must shift if we value sustainable development.

Scientists need the private sector to provide funding and a 'tunnel' for commercialization; the private sector needs scientists to develop products. This alliance, together with support from the government, is called the triple helix — a concept that has driven the world's economy since the Industrial Revolution. But is this concept still relevant?

Although some parts of the world have achieved a stable economy driven by scientific

advancement, around half of the world's population still lives in poverty. The people of these regions also face environmental threats, such as deforestation and its extended impact, on a daily basis. Those who are most vulnerable benefit from science the least.

There are scientists who want to transfer their knowledge to these people, but this has proved difficult. The failure of an experiment in the Solomon Islands to help indigenous people to exploit their local environment as 'ecosystem services' was attributed to a culture gap between scientists and local people. This claimed divide is often presented as a barrier to the transfer of science and technology.

Scientists must try harder to bridge this gap. Science is a fuel for economic development, but its influence must extend beyond the triple helix. That model simply uses science to exploit natural resources for economic gain. Given the need to mitigate the harmful environmental effects of this conversion, the model is no longer enough.

> **MITIGATION** MUST BE THE RESPONSIBILITY OF **EVERYONE** ON THE PLANET, NOT JUST **SCIENTISTS,** BUSINESSMEN AND POLICYMAKERS.

Mitigation must be the responsibility of everyone on the planet, not just scientists, businessmen and policymakers. Indigenous and local people should also be involved, especially those who call carbon sinks, such as tropical forests, home.

There are already examples of science reaching out. The residents of the Wanang Conservation Area in Papua New Guinea, for instance, have offered 1,000 hectares of their 10,000-hectare protected forest for research conducted by institutions such as the Smithsonian Tropical Research Institute's Center for Tropical Forest Science. In this zone, scientists and indigenous people collaborate to investigate the response of trees to climate change. Local people are trained then employed as field research assistants and have received compensation for the lease of their forest.

Meanwhile, a project supported by the US Agency for International Development is training local people in West Kalimantan, Indonesia, to be plant parataxonomists. The project was initiated by Campbell Webb, a plant evolutionary biologist and bioinformaticist at the Arnold Arboretum of Harvard University who is based in West Kalimantan. It is teaching local people to collect plant data in Gunung Palung National Park, an area of high biodiversity that faces the threat of deforestation.

The Paris talks should discuss the need for such initiatives to be copied and scaled up. For decades, the relationship between science, industry and government has been celebrated by all involved as a good thing. But not everybody benefits. Science might be able to pin the blame for the southeast Asia haze on Indonesian smallholders, but it has not yet given them — or others in their position — a way to help prevent it. ∎

Dyna Rochmyaningsih *is a freelance science journalist in Sumatra.*
e-mail: drochmya87@gmail.com

NANOMATERIALS

Droplets surf graphene waves

Tiny particles of liquid move quickly across thin layers of carbon by 'surfing' waves that ripple through the sheets.

Angelos Michaelides at University College London and his colleagues used computer simulations to investigate how liquids move across graphene — a layer of carbon one atom thick. Graphene has wave-like ripples that transport nanometre-scale droplets of water and oil, and even ice particles. This happens because the particles are attracted to the high density of carbon atoms in the wave trough. These nanodroplets move much more quickly on flexible layers of material such as graphene than on rigid materials like metal.

If validated by experiments, this mechanism could be used to control the delivery of water-soluble drugs on surfaces coated with a layered material, the authors say.
Nature Mater. **http://dx.doi. org/10.1038/nmat4449 (2015)**

PROSTHETICS

Bionic touch lights up neurons

A thin, flexible device can sense a wide range of pressures and produces signals that stimulate nerve cells in a dish.

Zhenan Bao of Stanford

PLANETARY SCIENCE

Pluto hosts wildly varying terrain

The first published findings from NASA's New Horizons mission to Pluto confirm that the dwarf planet has geological features that resemble those found on Mars and various moons in the Solar System.

NASA's spacecraft flew past Pluto in July, sending back reams of data that have been analysed by Alan Stern at the Southwest Research Institute in Boulder, Colorado, and his colleagues. Broad, bright plains on Pluto known as Sputnik Planum seem to be covered by nitrogen glaciers; these quickly erase craters made by crashing asteroids. Nearby lies the dark Cthulhu region, which is covered in craters that are thought to be up to 4 billion years old.

Pluto also hosts unique features, such as 'snakeskin' terrain that may have been sharpened into ridges over time as material froze and then sublimated away.
Science 350, **292 (2015)**

University in California and her collaborators embedded carbon nanotubes in a rubbery polymer and attached that material to a flexible circuit (**pictured** mounted on a robotic hand). The device mimicked the response of touch-sensitive nerve cells in the skin by emitting discrete electrical spikes of increasing frequency in response to applied pressure. The team converted the electronic signal into light that then stimulated genetically engineered, light-sensitive mouse neurons *in vitro*.

Such artificial skin could one day restore sensation for people wearing prostheses, the authors say.
Science 350, **313–316 (2015)**

ECOLOGY

Caffeine keeps bees coming back

Caffeine-infused nectar tricks honeybees into changing their foraging behaviour in ways that may benefit the plant.

Many plants produce the bitter-tasting caffeine to deter herbivores, but also rely on bees to spread their pollen for reproduction. To look at caffeine's effect on pollinators, Margaret Couvillon and her colleagues at the University of Sussex near Brighton, UK, monitored honeybees feeding from a sugar solution. They then compared the bees' behaviour to those feeding on the same solution but with caffeine added at

a concentration found in nectar. The caffeine-fuelled bees revisited the feeders more frequently than did the control bees, and they at least tripled the number of waggle dances they performed to recruit bees from the hive.

Because caffeine disguises a reduced sugar concentration, the nectar the bees take back to the hive might be sub-standard. That could mean that the colony would produce less honey, the authors predict.

Curr. Biol. http://dx.doi.org/10.1016/j.cub.2015.08.052 (2015)

ANIMAL BEHAVIOUR

Electric eels use shocks to sense

Electric eels send out strong zaps to track moving prey by their electrical conductivity, enabling the eels to strike with remarkable precision.

Electric eels (*Electrophorus electricus*; **pictured**) are known to use electricity to stun their prey, and have electrical sensors (pictured in pink). To see whether the high-voltage zaps have a sensory role, Kenneth Catania at Vanderbilt University in Nashville, Tennessee, presented the eels with a twitching fish in an insulated plastic bag and a conductive rod.

The eels reacted to the mechanical signals from the moving prey, producing a strong shock and striking in the direction of the fish. But they repositioned mid-strike, capturing and attempting to feed on the rod instead, even when it moved around quickly.

This sensory system is

NATURE COMMUN./CC-BY

similar to how some bats use echolocation, says Catania.
Nature Commun. 6, 8661 (2015)

AGROECOLOGY

Wild flowers are a pesticide source

Commonly used insecticides have been found on wild flowers as well as on crops.

Neonicotinoid pesticides applied to the seeds of some crops end up in the nectar and pollen of adult plants, so the chemicals are a suspected cause of the global decline in bee populations. Because most crops flower only briefly, it was unclear how bees could be exposed to enough pesticide to feel toxic effects. Now Cristina Botías and her colleagues at the University of Sussex in Brighton, UK, show that these chemicals are present in the pollen of wild flowers growing near fields where neonicotinoids were used.

The team measured neonicotinoid levels in pollen sampled from fields of oilseed rape (*Brassica napus*), nearby wild flowers and local beehives, and estimated that 97% of these compounds that were brought back to beehives originate from wild flowers.

The wild flowers had higher levels of insecticide in their pollen than crop plants did, and they bloom for much longer.
Environ. Sci. Technol. http://doi.org/8bk (2015)

MEDICAL TECHNOLOGY

Cheap MRI uses small magnets

A technique for magnetic resonance imaging (MRI) could provide fast brain scans at a fraction of the cost of conventional machines.

Most MRI scanners require large magnets to generate a strong enough magnetic field to penetrate soft tissue. A team led by Matthew Rosen at Harvard Medical School in Boston, Massachusetts, has demonstrated a way to capture an image using magnetic fields that are 450 times weaker than those used by current machines, and at one-twentieth of the cost.

The team engineered a radio-frequency coil that could pick up the faint radio signals generated as a result of the weak magnet and used data-collection techniques that speed up image reconstructions.

Although the resulting images have a lower resolution than do those from large MRI scanners, they can still reveal major abnormalities such as signs of traumatic brain injury or stroke, Rosen says.
Sci. Rep. 5, 15177 (2015)

EVOLUTION

Village-dog DNA hints at origins

DNA from free-roaming 'village dogs' shows greater genetic diversity than that of pure-bred dogs, and could help to settle debates about where dogs were domesticated.

Humans domesticated dogs from wolves more than 15,000 years ago, but researchers disagree about whether that happened in Europe, East Asia, the Middle East or elsewhere. A team led by Adam Boyko at Cornell University in Ithaca, New York, analysed the genomes of 549 free-breeding village dogs from around the world, as well as 4,676 pure-bred dogs belonging to 161 breeds. Genome-wide patterns of ancestry in the village dogs hint at a central Asian origin for domestic dogs, followed by population expansions in East Asia.

The researchers say, however, that more-extensive studies of DNA from diverse dogs are needed to pinpoint the origins of man's best friend.
Proc. Natl Acad. Sci. USA http://dx.doi.org/10.1073/pnas.1516215112 (2015)

CORRECTION

The print version of the Research Highlight 'Corals cope with acidified waters' (*Nature* **526,** 296–297; 2015) incorrectly stated that ocean water is being acidified when in fact it is becoming less alkaline; the online title was changed to reflect that. It also said coral-made fluid was less acidic than reef waters; in fact, the fluid had a higher pH. And it said that some corals can control the pH of surroundings, whereas they control their internal pH.

↪ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS
The news in brief

Marcy resigns

Astronomer Geoffrey Marcy resigned from the University of California, Berkeley, on 14 October, following revelations that he had violated his university's sexual-harassment policies. In response, the American Astronomical Society is updating its code of ethics to include guidelines and practices for dealing with misconduct. Marcy, a pioneer in the field of exoplanets, has also terminated his relationship with the Breakthrough Listen project to search for extraterrestrial intelligence, and been removed from an adjunct position at San Francisco State University. See page 483 for more.

EVENTS

New Ebola cases

The World Health Organization (WHO) reported two new cases of Ebola in Guinea on 16 October, ending a two-week period in which no new cases had been detected across West Africa. Contacts of both individuals will receive an experimental Ebola vaccine as part of an ongoing clinical trial. The WHO does not consider a region Ebola-free until 42 days have passed without a new case.

Harvard and China

Harvard University in Cambridge, Massachusetts, has unveiled a collaborative environmental research project in China. The US$3.75-million venture, announced on 15 October, will enable atmospheric scientist Michael McElroy to work with Chinese researchers on climate change, energy security and sustainable development. Based at the Harvard Center Shanghai, it will include studies in economics,

RALPH CLAUS GRIMM

Pollen-coated honeybee photo wins gold

This extreme close-up of a honeybee (*Apis mellifera*) eye covered in dandelion pollen grains won the annual Nikon Small World Photomicrography Competition. The contest showcases microscopic images captured worldwide by scientists, artists and others. Australian secondary-school teacher Ralph Grimm, a former beekeeper, snapped the photo, devoting four hours to mount and light the eye and focus the image. Second prize went to an image of a mouse colon colonized with human microbiota, and in third place was a picture of the trap of a humped bladderwort (*Utricularia gibba*), a freshwater carnivorous plant.

engineering, atmospheric science and environmental health related to sustainability. The collaboration is the first initiative of the Harvard Global Institute, a funding mechanism launched the same day, to encourage interdisciplinary collaborative research overseas.

Whaling fight

Australia's environment minister said on 19 October that the country is taking legal advice over Japanese plans to restart whaling in the Southern Ocean, claiming that Japan is attempting "to exclude itself from the International Court of Justice in matters relating to future whaling activities". The court declared Japan's whaling in the region illegal in 2014, ruling that it was not strictly for scientific purposes. Minister Greg Hunt said that Australia had met with the Japanese government to discuss the latter's apparent attempt to sidestep the court.
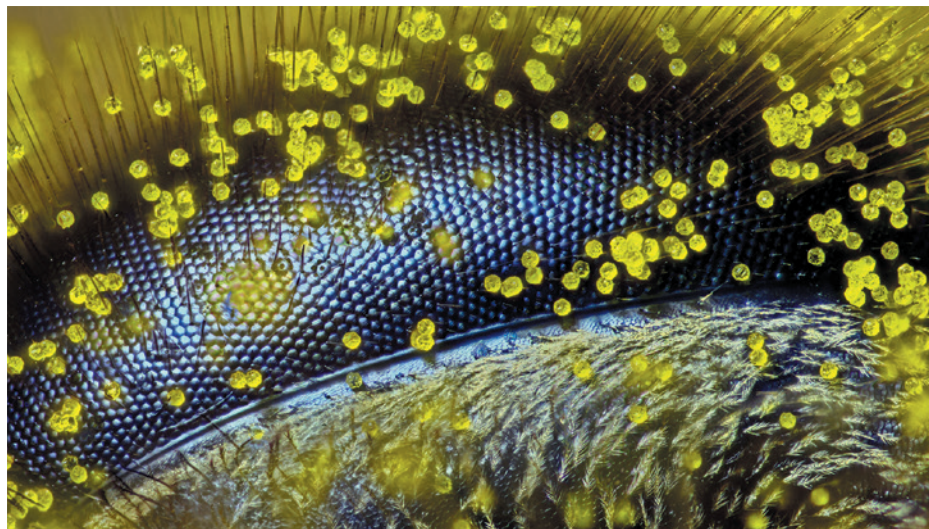
Oil-industry pledge

Officials from ten of the world's largest oil and gas companies, including BP and Shell, endorsed on 16 October the goal of limiting the increase in average global surface temperatures to 2 °C. While calling on governments to create "clear stable policy frameworks", the companies committed to increasing investments in clean-energy solutions, including energy efficiency and technologies that would allow carbon dioxide from industrial plants to be captured and sequestered underground. Released less than two months before the United Nations climate summit in Paris, the statement was met with scepticism from environmentalists, who say that the industry is still fighting meaningful climate regulations.

WHOI cyberattack

The Woods Hole Oceanographic Institution (WHOI) in Massachusetts told its staff on 13 October that it had been the target of a cyberattack. An investigation suggests that the attack originated in China. The institution does both classified and unclassified oceanographic research; no classified information was accessed during the breach, it says. The attack, which targeted data

and e-mail, began as early as February 2013 and was not detected until June 2015.

Canada leans left

The Liberal Party triumphed over the Conservatives in Canada's 19 October general election. Incoming prime minister Justin Trudeau pledged before the election to appoint a chief science officer to make government science "fully available to the public". The previous Conservative government attracted large-scale protests from researchers who accused its leader Stephen Harper of muzzling federal scientists and cutting research budgets. See go.nature. com/27d1td for more.

POLICY

Halt to Arctic oil

The US Department of the Interior on 16 October cancelled a pair of oil and gas lease sales in the Arctic, citing low oil prices and weak industry interest. The decision comes after the oil company Shell suspended its Chukchi Sea Arctic drilling programme in September after finding less oil and gas than expected. The lease sales had been planned for 2016 in the Chukchi Sea and 2017 in the Beaufort Sea (**pictured**). The department also denied requests from Shell and Norwegian oil company

Statoil to put their existing Arctic leases on hold and resume them at a later date.

Oversight overhaul

The US National Institutes of Health (NIH) may scale back its review of research involving human gene therapy. The agency's Recombinant DNA Advisory Committee reviews all research protocols involving gene transfer into humans, but on 16 October, the NIH proposed that such research be reviewed only if requested by local ethics committees. A 2013 report from the US Institute of Medicine argued that the current level of oversight is no longer needed. The public is invited to comment on the proposal until 30 November.

Emissions action

The White House announced a series of executive actions and voluntary industry commitments on 15 October to reduce emissions of hydrofluorocarbons (HFCs),

a group of potent greenhouse gases that are commonly used as refrigerants. The US Environmental Protection Agency said that it will pursue new rules governing the use and management of HFCs, and the Department of Defense announced plans to use alternative chemicals at some of its facilities and on ships. Combined with earlier announcements, the commitments aim to reduce global greenhouse-gas emissions by the equivalent of more than 1 billion tonnes of carbon dioxide by 2025.

Wellcome boost

The Wellcome Trust, Britain's largest biomedical-research charity, announced plans on 21 October to spend £5 billion (US$7.7 billion) over the next five years. The organization has disbursed £6 billion over the past 10 years, including £728 million in 2014. Jeremy Farrar, the trust's director, says that Wellcome "will do
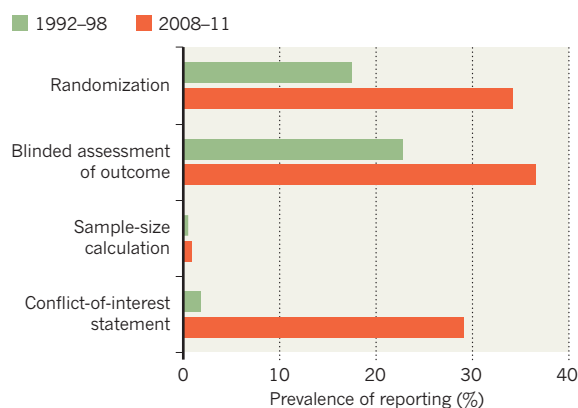
more of what we're already doing", and identified other areas such as combating drug-resistant infections and research that involves mining medical records. The splurge is driven by the performance of Wellcome's £18 billion endowment fund, says Farrar. From 2013 to 2014, its investment assets grew by about 10%, to £18 billion.

CORRECTION

The story 'Telescope start' (*Nature* **526,** 298; 2015) stated that the 23-metre telescope being built in the Canary Islands would form part of the Cherenkov Telescope Array. However, the telescope is a prototype and will not necessarily become part of the array, which has yet to be finally approved or funded.

↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Drug testing in animals is at substantial risk of bias because of poor study design, suggests an analysis of thousands of papers (M. R. Macleod *et al. PLoS Biol.* http://doi.org/8cf; 2015). Many publications do not mention bias-avoiding methods. These include randomizing animals' assignment to treatment or control arms; calculating the sample size necessary for a statistically robust result; and 'blinding' researchers as to which animals were assigned which treatment. See go.nature. com/j7ipin for more.

ANIMAL STUDIES POORLY DESIGNED
Studies testing drugs in animals rarely report the use of basic methods to avoid biased conclusions.

■ 1992–98   ■ 2008–11

Category	
Randomization	
Blinded assessment of outcome	
Sample-size calculation	
Conflict-of-interest statement	

Prevalence of reporting (%)
0   10   20   30   40

NEWS IN FOCUS

TECHNOLOGY Health-care ventures draw biomedical scientists to Silicon Valley **p.484** | **RUSSIA** Expanded secrets law requires security review of all science papers **p.486** | **CONSERVATION** US and Cuba cooperate on shark-management plan **p.488** | **TIME USE** The lab that uses 850,000 diaries to work out where your time goes **p.492**

Exoplanet-hunter pioneer Geoffrey Marcy has resigned from his job and some of his research projects.

SOCIETY

Astronomy moves to end harassment

US researchers unite against abuse in the workplace.

BY ALEXANDRA WITZE

Twenty years ago this month, Geoffrey Marcy narrowly missed out on becoming the first astronomer to find a planet orbiting a distant star. Beaten to the discovery by a pair of Swiss scientists, Marcy and his colleagues went on to rack up an astonishing list of other extrasolar planet sightings, from the first multiplanet system around a Sun-like star to the first Neptune-sized exoplanet. It was the kind of career that triggered talk of a Nobel prize.

On 14 October, Marcy resigned from the University of California, Berkeley, in the wake of a sexual-harassment scandal that involved multiple students over many years. He leaves a field that has expanded far beyond his early influence — and many researchers who hope that the harassment revelations will lead to improvements in working conditions for women in astronomy.

Having such a prominent researcher involved in a high-profile harassment case may prompt more scientists to recognize the deep-seated problem, says Julianne Dalcanton, an astronomer at the University of Washington in Seattle. "Perhaps moving forward, more people will be part of the solution," she says. Already, astronomy departments at many universities are starting to hold open discussions about how to prevent abuses on their campuses.

"The damage he has caused and the culture which enabled it to happen still need to be addressed," adds Laura Lopez, an astronomer at the Ohio State University in Columbus.

Statistics paint a grim picture for US women in astronomy. Just 14% of full professors in the field at US universities are women, according to a 2013 survey by the American Astronomical Society (AAS) Committee on the Status of Women in Astronomy. Studies suggest that sexual harassment is pervasive in academia. In an April 2015 survey of students, faculty members, staff and alumni in Marcy's department at Berkeley, more than one-third of 45 women reported some form of sexual or gendered discomfort brought on by the actions of other members of the department.

SCANDAL ERUPTS

Prompted by formal complaints, Berkeley investigated Marcy and concluded in June that he had violated campus sexual-harassment policies in incidents involving students between 2001 and 2010. (Marcy became a professor at the university in 1999.) The revelations became public in a 9 October article on BuzzFeed News. Berkeley administrators said that because they could not unilaterally discipline a faculty member, they had reached an agreement with Marcy in which he would be stripped of faculty career protections and be subject to sanctions or dismissal if he violated policies again.

Astronomy faculty members and students at Berkeley protested against the university's response, and Marcy resigned. San Francisco State University, where he worked before Berkeley and had retained an adjunct position, terminated its relationship with him.

The fate of Marcy's research projects remains unclear. That includes his work helping to lead NASA's planet-hunting Kepler mission, and with the Automated Planet Finder, a robotic 2.4-metre telescope at Lick Observatory in northern California that searches for rocky planets. Marcy has resigned as a principal investigator of Breakthrough Listen, a US$100-million project announced in July to accelerate the search for signs of intelligent life in the Universe. The mission will continue to be overseen by Berkeley astronomers Andrew Siemion and Dan Werthimer, among others. ▶

▶ Although Marcy was a pioneer in exoplanet research, the field has grown far beyond him, says Mercedes López-Morales, an astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. She hopes that the quick and nearly universal condemnation of Marcy's actions will encourage young researchers, particularly women, to pursue exoplanet research. "Marcy represents the exception, not the rule, in our field," says López-Morales.

"We should all take a close look at our own institutions and professional networks and ask what we might do differently."

Among those swift responses was a statement from the AAS. The Marcy case "offers an important opportunity for all of us to discuss, within our groups and institutions, what responsibilities we have as professionals and how we can ensure that everyone in our profession is afforded a safe, supportive workplace", it reads. AAS president Meg Urry, an astronomer at Yale University in New Haven, Connecticut, is a long-time advocate for improving working conditions for women. After the Marcy revelations, Urry set up a task force to develop procedures and sanctions related to misconduct, for inclusion in the society's code of ethics.

"We should all take a close look at our own institutions and professional networks and ask what we might do differently," says Heather Knutson, an exoplanet researcher at the California Institute of Technology in Pasadena.

Compared with other fields of science and other countries, US astronomy has been relatively progressive in tackling workplace issues for women and other minorities. The AAS Committee on the Status of Women in Astronomy runs a website with discussion and specific advice on topics such as bullying. Volunteers have also started a programme called Astronomy Allies, which serves as a buddy system to walk people home from astronomy-related parties and conference events.

Other research areas should also pay attention, Lopez says. "Sexual harassment is a problem endemic to all fields in academia," she says, "and Marcy's case should serve as a reality check for everyone, not just for astronomers." ∎

Google provides colourful bikes for employees to ride around its California campus.

BIOTECHNOLOGY

Tech titans lure life–sciences elite

As Google and others turn to health care, biomedical luminaries flock to Silicon Valley.

BY ERIKA CHECK HAYDEN

Free tasty food, brightly coloured bicycles and high salaries are well-known hallmarks of the Googleplex — Google's famed headquarters in Mountain View, California. But it was not these perks that led cardiologist Jessica Mega to pause her thriving academic career at Harvard Medical School to become the chief medical officer of the company's life-sciences team. She was lured by the ambitions of the effort, soon to be incorporated under Google's parent firm Alphabet. Nurtured by Google's expertise in data analytics and engineering, the biology team is expected to create miniaturized electronic devices and to use these and other means to collect and analyse more health data, more continuously, than is possible today.

"What I find compelling is the immersion of people with strong technology backgrounds — hardware and software engineers — sitting next to people like myself," says Mega. "The impact feels very, very large."

Mega's decision to move in March to Google was one in a string of announcements by top-flight scientists and physicians who are enlisting in the mission, and pioneering

JUSTIN SULLIVAN/GETTY

MORE ONLINE

TOP STORY
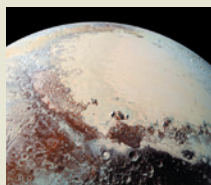
Pluto's geology is unlike any other in the Solar System go.nature.com/7pmbnp

MORE NEWS

● Vast cosmic voids merge like soap bubbles go.nature.com/fjdgoc
● First report of sexually transmitted Ebola go.nature.com/svhmtb
● Teeth from China reveal early human trek out of Africa go.nature.com/mevtgq

NATURE PODCAST

A dying solar system like ours; how heat affects economies; and electricity-eating bacteria nature.com/nature/podcast

NASA/JHUAPL/SWRI

a new type of career path in the process. Although academic researchers from fields such as computer science and engineering have led innovative Google projects (such as the Internet-connected eyewear known as Glass), Google and other technology companies are increasingly recruiting life scientists as Silicon Valley broadens its reach into health care. "I have a feeling we're going to see a lot more recruitment of leading lights," says Eric Topol, director of the Scripps Translational Science Institute in La Jolla, California.

In September, Thomas Insel, director of the US National Institute of Mental Health in Bethesda, Maryland, announced that he would soon be joining Google's life-sciences company to help develop ways to apply technology in mental health. And last year, molecular biologist Cynthia Kenyon, a leader in ageing research at the University of California, San Francisco, joined the Google-backed biotech company Calico in San Francisco, California.

Cardiologist Euan Ashley of Stanford University, which sits in the thick of Silicon Valley, says that academic data scientists are constantly tempted by the companies that await them just off campus. "They're being continuously recruited away," he says. "We're in competition with Google and other tech companies, and generally they can pay a lot more than Stanford can."

But money is not the only lure. Silicon Valley offers strong technology resources that are hard to access in academia, Topol says, as well as the opportunity to pursue goals that are difficult to reach for in academia, where scientists are not typically rewarded for pursuing real-world applications. "The resources are exponentially greater than what you can get through academic circles. And the metrics are different: instead of publications, it's just, 'Get stuff done'," he says.

Getting stuff done was foremost in the mind of electrical engineer Brian Otis when he left his tenured position at the University of Washington in Seattle in 2012 to work for Google. He went there to work on a 'smart' contact lens for people with diabetes that measures the level of glucose in tears. When the project began, it faced two big questions: first, could the electronics needed to make a functional wireless glucose sensor be embedded in a wearable contact lens? And second, would it provide the relevant measurements of glucose levels? The motivation and means to answer those unknowns was a powerful incentive, Otis says. He recalls thinking: "If I come into Google life with these questions, I have the entire runway and resources to answer these two questions."

The project was successful; drug giant Novartis licensed the contact-lens technology last year and Otis is now director of the Google life-sciences team's hardware and medical-device development. "To go all the way from foundational first principles to execution of vision was the initial draw, and that's what has continued to keep me here," he says.

Apple, too, has entered the health-care game. In March, it debuted ResearchKit, a framework through which researchers can write apps that collect data from patients' mobile phones. And in April, IBM launched IBM Watson Health and the Watson Health Cloud, services that use the company's cognitive computing technology to process large amounts of health data from diverse sources. The service could help physicians to manage patients' health by streaming data from personal electronic devices, or enable drug companies to manage clinical trials more efficiently with cloud computing. Intel, meanwhile, is developing cloud-computing services to provide more personalized cancer care; and Facebook, Microsoft and Amazon are all also getting involved.

But Google's approach sets it apart: the company expends more resources on potential health applications and is exploring in more directions than others are. Observers estimate that Google puts more than a billion dollars per year into life-sciences research, although the company says that it does not break down its spending in that way.

"The resources are exponentially greater than what you can get through academic circles."

Google's life-sciences team is working on a range of projects that involve developing new ways of monitoring health. As well as the smart contact-lens project, there is the Baseline Study, which aims to collect large amounts of data about people to better quantify health and disease, with the goal of earlier and more-effective preventive care. The company also funds a huge array of external collaborations with academics. Google Genomics, for instance, is studying the application of cloud computing to genomics, and Calico has signed a slew of collaborations with companies and academic institutes.

"They're reaching out to academia in a way that biotechnology companies often don't," says cell and molecular biologist Judith Campisi of the Buck Institute for Research on Aging in Novato, California. That enables scientists to collaborate with Google instead of joining it wholesale.

"For some academics, joining a technology company would be an exciting new opportunity," says physician Steven Hyman of the Broad Institute of MIT and Harvard in Cambridge, Massachusetts. But it is "not a likely destination for those interested in mitigating risk," he says. "After all, the life-science goals of the Googles, Apples and Microsofts of the world are likely to change in the near term as the companies explore an area that is new to them." ∎

Neutrino probe is key priority for US physics

Nuclear-science wish list also includes particle collider.

BY DAVIDE CASTELVECCHI

Two large science experiments head a wish list drawn up by US nuclear physicists for the next decade: a quest to uncover the nature of neutrinos and a particle collider to study the forces that bind quarks.

The big-ticket items, each of which would cost hundreds of millions of dollars, are among the top priorities highlighted by the Nuclear Science Advisory Committee (NSAC) on 15 October. Every 5–7 years, this panel of high-level nuclear physicists presents a long-term plan to the US Department of Energy and National Science Foundation, after consulting the US nuclear-physics community.

The agenda assumes that US funding for nuclear science will increase by 1.6% per year above inflation — a realistic scenario, says NSAC chair Donald Geesaman, a physicist at Argonne National Laboratory in Illinois. "We have exciting science to do, and we are not asking for large increases," he says.

The neutrino experiment, construction of which could begin by the end of the decade, would search for a theorized rare form of radioactive decay in which two identical neutrinos annihilate one another — an event that would imply that neutrinos are their own anti-particles. It could provide a way to measure the tiny mass of neutrinos and help to explain why the Universe has lots of matter but almost no antimatter.

Experiments around the world using materials such as liquid xenon have failed to detect the event, known as neutrinoless double β decay. But the NSAC report says that an experiment using a tonne or more of material — about ten times more than any previous attempt — could either find or rule out the phenomenon.

Another priority, on which *Nature* reported in May (see *Nature* **521,** 272; 2015), is a particle accelerator that would collide electrons with protons or heavy ions to investigate gluons, which carry the force that binds quarks. But construction would have to wait until the 2020s because NSAC's top priority is to complete and maintain existing facilities, such as the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory in Upton, New York. RHIC faced closure two years ago, but an improved budgetary position means it can now be sustained into the next decade. ∎
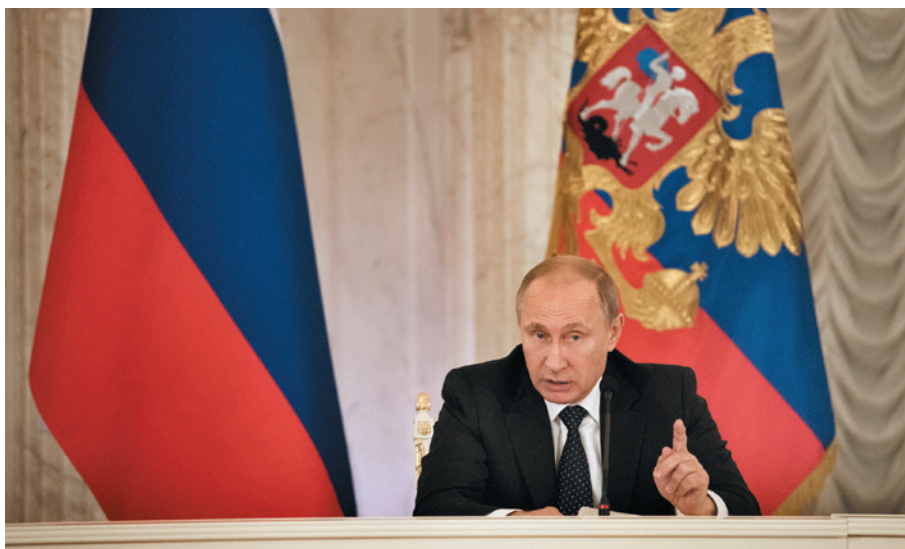
Russian President Vladimir Putin has declared certain types of civilian research classified.

RUSSIA

Secret service to vet manuscripts

Moscow biology department requires research papers to be approved to comply with law on state secrets.

BY QUIRIN SCHIERMEIER

A biology institute at Russia's largest and most prestigious university has instructed its scientists to get all research manuscripts approved by the security service before submitting them to conferences or journals.

The instructions, which come in response to an amended law on state secrets, appear in minutes from a meeting held on 5 October at the A. N. Belozersky Institute of Physico-Chemical Biology at Lomonosov Moscow State University (MSU).

The Russian government says that the amendment is not designed to restrict the publication of basic, non-military research. But scientists say that they believe institutes across the country are issuing similar orders.

"This is a return to Soviet times when in order to send a paper to an international journal, we had to get a permission specifying that the result is not new and important and hence may be published abroad," says Mikhail Gelfand, a bioinformatician at MSU.

In 1993, the government passed a law obliging scientists in Russia to get permission from the Federal Security Service (FSB) before publishing results that might have military or industrial significance. This mainly covered work that related to building weapons, including nuclear, biological and chemical ones.

However, in May, President Vladimir Putin used a decree to expand the scope of the law to include any science that can be used to develop vaguely defined "new products". The amendment was part of a broader crackdown that included declaring the deaths and wounding of soldiers during peacetime a secret; this was prompted by accusations that Russian soldiers are involved in conflict in Ukraine.

Since then, rumours have emerged that Russian universities and institutes are demanding that manuscripts be approved before submission to comply with the amendment. The minutes from the Belozersky Institute meeting confirm this. "Be reminded that current legislation obliges scientists to get approval prior to publication of any article and conference talk or poster," they say. They note that the rules apply to any publication or conference, foreign or national, and to all staff "without exception".

Scientists will need to seek permission from the university's First Department — a branch of the FSB that exists at all Russian universities and research institutes, says Viacheslav Shuper, a geographer at the Russian Academy of Sciences in Moscow and MSU. He says that MSU geographers have been given similar instructions.

The minutes tell scientists to seek permission "despite the obvious absurdity of the whole situation". Vladimir Skulachev, director of the Belozersky Institute, did not respond to *Nature*'s queries as to how the changes might affect research in his department.

Shuper and other academics say researchers across Russia have complained that their institutes are also asking for manuscript approval. "Many scientists in Russia don't dare to speak openly," says Shuper. "But I know that many are very unhappy about the degradation of their academic freedom."

Letting bureaucrats decide whether any piece of science is a state secret is not just nerve-wracking, but also burdensome, he says. For example, at some institutes, scientists who have written papers in English for foreign publication are obliged to translate them into Russian for the sake of the security service.

The changes are also bad for science, says Fyodor Kondrashov, a Russian biologist at the Centre for Genomic Regulation in Barcelona, Spain. "The problem is that it appears that all scientific output is being treated as potentially classified," he says. "This creates an unhealthy research climate with some scientists preferring not to share information — not to give a talk at a conference abroad, for example. I fear that the authorities will choose to apply this law selectively against their critics."

Sergey Salikhov, director of the Russian science ministry's science and technology department, told *Nature* that the government does not intend the amendment to restrict the publication of basic research. He says that it is not ordering universities or security services to proactively enforce the law over civilian research.

But the amendment leaves interpretation to the security services and science administrators, who tend to be over-zealous, says Gelfand. "Basically, anything new and potentially useful can now be interpreted to be a state secret," says Konstantin Severinov, a molecular biologist with the Skolkovo Institute of Science and Technology, who graduated from MSU.

The demand for approval runs counter to government efforts to strengthen and internationalize Russian science, says Severinov. The government aims to see 5 of the country's universities enter the top 100 in the world rankings by 2020, and is keen to attract leading foreign scientists to Russia.

Gelfand says that he will not comply with the rules imposed by his institute, and he encourages others to follow suit. "A sad sign of overall deterioration here is that many are sheepishly following any absurdity instilled by the bureaucrats," he says. "I am going to ignore it and hope that a sufficient number of colleagues would do the same." ■ **SEE EDITORIAL P.475**

> *"Anything new and potentially useful can now be interpreted to be a state secret."*

GENETIC MEDICINE

Gene therapy sees early success against progressive blindness

Treatments for inherited eye diseases show promise in clinical trials, but worries linger over how long the beneficial effects will last.

BY HEIDI LEDFORD

Ophthalmologist Eric Pierce is no stranger to difficult conversations. During his years at the Boston hospital Massachusetts Eye and Ear, he has counselled both adults and the parents of young children who have been newly diagnosed with genetic retinal diseases that will ultimately leave them blind.

But progress against one such disease has led Pierce to change how he presents his diagnosis. "We're on the threshold of a new era," he now tells anxious parents. "I do believe there will be a therapy for your child so that they won't experience the full course of this disease."

Pierce's optimism is grounded in early data from tests of gene therapy in animals and humans. On 12 October, researchers reported success with using gene therapy in dogs against a form of retinitis pigmentosa[1], a genetic disease that causes light-sensitive photoreceptor cells to degenerate over the course of years. The results unexpectedly showed that the approach worked well even in mature dogs that had already lost some photoreceptor cells, a sign that the strategy might also work in humans, who have often reached that stage well before diagnosis.

And on 10 October at the Retina Society annual scientific meeting in Paris, a biotechnology company presented encouraging data from a trial in humans. The company found that its gene therapy for a degenerative eye disease caused by a mutation in the *RPE65* gene improved sensitivity to light in all 21 treated patients. Although other research groups using the same approach have seen some reversal of similar gains, the company, Spark Therapeutics of Philadelphia, Pennsylvania, plans to apply to the US Food and Drug Administration for regulatory approval of its therapy in 2016. If the treatment is approved, the company could be the first to bring a gene therapy to market in the United States.

"This is definitely a time of great promise," says Stephen Rose, chief research officer at the Foundation Fighting Blindness in Columbia, Maryland. "It moves the whole field of gene therapy forward."

Gene therapy has endured a bumpy road. After years of promising advances, the field almost came to a screeching halt in 1999. A death in a trial of gene therapy to treat an inherited metabolic disorder caused a scandal and sowed fears about the technology's safety. But an ardent few continued in the face of widespread scepticism and limited funding. The most notable success was a treatment for the genetic immune deficiency disease X-SCID, although it caused leukaemia in some patients.

It was during this time that some gene-therapy researchers began to see a glimmer of promise in treating eye disorders. The eye is partially shielded from the immune system, reducing the likelihood of an immune attack on the virus used to introduce the genes. (Such an immune response was blamed for the 1999 death.) The eye is also relatively easy to access, allowing surgeons to inject the virus near to the cells in which the gene is needed. Because more than 200 genes are associated with retinal disorders, the opportunity for genetic correction was clear.

Researchers began with mutations in *RPE65* that are associated with one type of vision loss. The enzyme encoded by *RPE65* is crucial for converting light into electrical signals that travel to the brain, and for sustaining the eye's photoreceptors. Without a functioning enzyme, the photoreceptors gradually degrade, progressively crippling vision. The researchers hoped to halt this process by using a virus to shuttle a functional *RPE65* gene into the eye.

In 2007, three teams launched the first clinical trials aiming to do just that, and included a team that would go on to found Spark in 2013. Positive results[2-4] published in 2008 rejuvenated interest in gene therapy, says Luk Vandenberghe, a virologist who studies gene therapy at Harvard Medical School in Boston. "They truly validated the concept of gene therapy that people had been pursuing for decades," he says. "The field has really turned around."

But earlier this year, two of those three teams announced setbacks. They reported[5,6] that the effects were waning in some patients as early as one year after treatment. ▶

VISION FOR THE FUTURE
Broader reach for gene therapy

Although gene therapy is showing promise against vision loss caused by mutations in the gene *RPE65*, such mutations account for less than 2% of the total burden of inherited diseases that cause retinal degeneration. Hundreds of genes have been implicated in such disorders; tackling them individually would require laborious research, clinical testing and regulatory approval for each one. Instead, researchers are seeking ways of using gene therapy to treat a larger set of patients.

Some are looking for ways to protect neurons in the eye from degeneration, regardless of which gene is involved in the process. Such 'neuroprotective' approaches under consideration include inducing cells to express a protein called RdCVF, which protects the cone cells in the eye that enable colour vision.

At Harvard Medical School in Boston, Massachusetts, Connie Cepko's laboratory is testing the effects of inducing the expression of a gene called *NRF2*, which activates antioxidant responses, to see whether those defences could protect photoreceptors from damage.

Others are taking a second look at GDNF, a neuroprotective protein that has been explored for its possible use as a treatment for Parkinson's disease and that may also protect photoreceptors in the eye.

Researchers at GenSight Biologics in Paris and at RetroSense Therapeutics in Ann Arbor, Michigan, are taking a different tack. They hope to replace damaged photoreceptors by inducing the eye's retinal cells to express light-sensitive proteins called channelrhodopsins. But channelrhodopsins are less sensitive to light than the eye's natural photoreceptors, so GenSight is also developing special goggles that patients would wear to amplify the light signals reaching the eye. **H.L.**

At Spark, chief scientific officer Kathy High says that her team has yet to see any decline in its patients even eight years after treatment. She notes that subtle differences in the protocol might have given Spark's treatment an edge. The virus that Spark engineered may have expressed *RPE65* at particularly high levels, she notes, and the company also adds a surfactant molecule when injecting the virus to prevent it from sticking to the needle during injection.

But vision scientist Artur Cideciyan of the University of Pennsylvania in Philadelphia, who works with one of the teams that reported a decline in gains after gene therapy, is still not convinced that Spark's results will endure. He says that Spark has not yet announced data as detailed as those that the other teams used to measure the growth — and then decline — in their patients' visual fields.

Even so, the diminishing effect in human trials need not indicate a fundamental flaw in the approach — or in gene therapy as a whole, says Vandenberghe (see 'Broader reach for gene therapy'). "All the tweaks haven't been fully worked out," he says.

Pierce, as a clinician, considers even tentative progress a huge achievement. He recalls the time when the only support that he could offer some of his patients was to recommend dietary supplements that might slow the disease. "Years of efficacy in a chronic degenerative disease is a huge success," he says. "And to have some optimism in the conversation is fantastic." ■

1. Beltran, W. A. *et al. Proc. Natl Acad. Sci. USA* http://dx.doi.org/10.1073/pnas.1509914112 (2015).
2. Hauswirth, W. W. *et al. Hum. Gene Ther.* **19**, 979–990 (2008).
3. Maguire, A. M. *et al. N. Engl. J. Med.* **358**, 2240–2248 (2008).
4. Bainbridge, J. W. B. *et al. N. Engl. J. Med.* **358**, 2231–2239 (2008).
5. Jacobson, S. G. *et al. N. Engl. J. Med.* **372**, 1920–1926 (2015).
6. Wright, A. F. *N. Engl. J. Med.* **372**, 1954–1955 (2015).

FISHERIES

Cuba forges links with United States to save sharks

Improved diplomatic relations feed a budding environmental partnership.

BY JEFF TOLLEFSON

Cuba is surrounded by sharks. Fishermen catch them, residents eat them and, increasingly, tourists are coming to see them. Now the island nation is gearing up to manage them, and its efforts are bolstering a nascent environmental partnership with the United States.

"It's a big step forward for Cuba and the region," says Jorge Angulo-Valdés, head of the Marine Conservation Group at the University of Havana's Center for Marine Research and a visiting professor at the University of Florida in Gainesville. "It's time for us to get together, identify common goals in resource management and make them work."

On 21 October, Cuba plans to release a management plan that will lay the groundwork for research and, eventually, regulations to protect extensive but largely undocumented shark and ray populations. Roughly half of the 100 species of shark resident in the Caribbean Sea and Gulf of Mexico have been seen in Cuban waters, including some — such as the whitetip (*Carcharhinus longimanus*) and longfin mako (*Isurus paucus*) — that have experienced sharp declines elsewhere. The Cuban government has consulted with environmentalists and academics from the United States and other countries in developing the plan.

"Cuba is a kind of biodiversity epicentre for sharks," says Robert Hueter, director of the Center for Shark Research at the Mote Marine Laboratory and Aquarium in Sarasota, Florida, who is one of those working with the Cuban

The Caribbean reef shark (*Carcharhinus perezii*) is one of many species that can be seen in Cuban waters.

PETE OXFORD/MINDEN PICTURES/GETTY

scientists. "The science is not at a level yet to do rigorous stock estimates, but we are moving in that direction with this plan."

Most of what is known about Cuba's shark populations has come from the fishing industry, which often captures sharks as by-products of its regular operations. The Cuban government has already established marine protected areas along 20% of its coastline and is planning to expand that network within the 70,000 square kilometres of its coastal fishery. It has also begun to regulate the equipment used in fishing, and is

looking to establish catch limits for various fish species, including sharks.

Both US and Cuban scientists say that the collaboration is helping to pave the way for more formal cooperation now that the two cold-war foes have re-established political relations. In April, the US National Oceanic and Atmospheric Administration (NOAA) sent a research vessel on a cruise around the island with Cuban scientists. And on 5 October, US secretary of state John Kerry and Cuban officials announced at an oceans conference in Chile that the two

nations were finalizing plans to cooperate on research, education and management in marine protected areas. The agreement could be finalized as early as next month, says Billy Causey, regional director for NOAA's Office of National Marine Sanctuaries in Key West, Florida.

POLITICAL IMPETUS

US environmentalists began pushing the idea of cooperation with Cuba on marine conservation after the 2008 election of President Barack Obama, who pledged during the campaign to engage with Cuba. The first signs of real progress came in September 2009, says Daniel Whittle, who heads the Cuba programme for the Environmental Defense Fund (EDF), an environmental group based in New York City. Then, the United States allowed four Cuban scientists, three of whom were marine and coastal researchers, to attend a series of meetings in the country. And in November last year, Angulo-Valdés was part of a cadre of Cuban scientists that visited the state department and several members of Congress. A month later, Obama ordered the restoration of diplomatic ties with Cuba.

"It's slowly beginning to change," says Whittle, referring to links between the nations. "That's why the announcement in Chile was so significant: finally the two governments publicly acknowledged that they are in fact working directly together on environmental issues."

The EDF and other conservation groups have been trying to build cooperation between Cuba, Mexico and the United States within the Gulf of Mexico. NOAA's April cruise, which focused on tallying the larvae of bluefin tuna (*Thunnus thynnus*) in Cuban and Mexican waters, marked the first formal government engagement on that front since Obama's December announcement, Causey says.

> *"Finally the two governments publicly acknowledged that they are in fact working directly together on environmental issues."*

The main question facing the shark-management plan is whether the Cuban government will be able to mobilize enough money to implement it. The EDF and other groups have been raising funds to pay for some of the initial work on the plan, including training fishing crews to identify and report the sharks that they catch. But scientists need to conduct population surveys that are independent of those done by commercial fisheries, and Cuban research institutions are already stretched thin.

The country has only two operational research vessels, and scarce resources to equip and operate them. The kind of tags needed to track shark movements through satellites can cost US$2,500 each. So far, Cuba has tagged just four sharks with such devices.

"We have to see how the government implements the plan, and how they get around the funding problem," Angulo-Valdés says. "It's going to be a challenge." ∎

Hunting the Godzilla El Niño

As a massive El Niño warming builds in the equatorial Pacific Ocean, researchers hope to make the most of their chance to study this havoc-wreaking phenomenon.

By Quirin Schiermeier

The tropical Pacific seemed out of sorts this August, as oceanographer Kelvin Richards and his team cruised along the equator east of the Marshall Islands. Six tropical cyclones had barrelled across the ocean in the previous month, and more were spinning up as Richards' research expedition got under way. The sea surface across the region was abnormally warm, with water temperatures at least 1 °C higher than expected. And when the oceanographers peered below the surface, they found signs of intense turbulence extending hundreds of metres down.

The team had found itself cruising through a spectacular El Niño warming event — one that may become the strongest ever recorded. Big El Niños can turn climate conditions in the Pacific upside down and disrupt weather around the globe. The impacts of this one have already been felt. Indonesia has suffered through a withering drought that has intensified fires in forests and agricultural land, and Pacific corals are experiencing one of the worst bleaching events on record. Peru has declared a state of emergency in some regions in expectation of flooding, and farmers in Australia have been put on alert for expected drought.

The last time a major El Niño developed, in 1997–98, extreme weather and flooding killed thousands and left a quarter of a billion people in Asia homeless. It also helped to jack up global temperatures to a point never recorded before.

For Richards, a researcher at the University of Hawaii at Manoa, the arrival of the latest El Niño turned out to be good timing. Such warmings develop only once or twice a decade, following no regular schedule, so researchers are eager to learn how to predict when an El Niño will hit and how powerful it will grow. That means that they must keep close tabs on the atmosphere and ocean, from the surface waters to the cooler layers hundreds of metres below. But getting the necessary data can be difficult. It takes years to plan research cruises, so it is hard to get a ship into the heart of the Pacific in time to study an unpredictable event. When Richards applied for ship time back in 2012, he had no idea that his trip would happen right as a warming episode was gathering strength. "It just happened to coincide with our expedition, and we gladly took the opportunity," he says.

Most oceanographers are not lucky enough to be out at sea this year, but they are taking advantage of their colleagues' data, as well as information flowing in from research buoys and other sources. One key question that they want to answer is why every El Niño behaves differently. "El Niños are not made from a cookie cutter," says Michael McPhaden, an oceanographer with the US National Oceanic and Atmospheric Administration (NOAA) in Seattle, Washington. The strength and impact of each El Niño seem to depend in part on which region of the Pacific warms up first, but predicting the pattern of temperature anomalies is tough. "We would really like to better understand what's causing the diversity, and how far in advance it might be possible to predict what type of event we need to prepare for," says McPhaden. That would help forecasters to give warning of coming droughts and floods months before they hit.

BAIT AND SWITCH

The current El Niño is a glaring reminder of how much scientists need to learn. When it first started to take shape in 2014, it developed like many others. There was a weakening in

The RV *Falkor* has collected data that will aid studies of El Niño.

the easterly trade winds that normally flow from South America towards Asia, carrying heat and moisture to the western part of the basin. This allowed warmth to spread eastwards, and researchers expected the pattern to be reinforced by westerly wind bursts that would help to push warm water eastwards (see 'Unruly ocean'). When enough warm water accumulates off the coast of South America, it prevents the normal upwelling of cool, nutrient-rich water from deeper layers. That, in turn, alters fish populations and typically ruins the anchovy harvest off the coast of Peru.

But in 2014, the warming along the equator was less pronounced than in most El Niño years, and the westerly wind bursts did not appear as expected. By mid-year, the anticipated El Niño had completely vanished.

What had stopped the show, and why the Pacific warming spectacularly resurfaced 12 months later, are questions that are puzzling ocean researchers and meteorologists. The mysteriously reborn El Niño is a fantastic opportunity for researchers to combine observations and models to find out what has happened, and perhaps to improve forecasting systems, says Axel Timmermann, an oceanographer at the University of Hawaii.

One possible explanation, he says, is that the expected westerly wind bursts came too early last year, so they did not pile up enough warm water in the eastern Pacific to inhibit upwelling. That would have stopped El Niño in its tracks. But there is also a chance that an overlooked mechanism enabled cool water from deep layers to reach the surface. Or it might simply be that the erratic nature of the El Niño Southern Oscillation (ENSO) — the irregular sequence of warm El Niño and cold 'La Niña' phases — is down to the randomness of the weather.

To test these hypotheses, researchers will need many forms of data, including measurements of ocean temperature over time, upwelling rates, water density and the strength of currents. And it will be important to compare El Niño years to neutral years and times when La Niña appears, as well as years when an event seems to be looming and then fails to materialize, says Matthew England, a climate scientist at the University of New South Wales in Sydney, Australia.

The problem is made even more difficult because ENSO behaviour may be shifting as a result of climate change. Warmer surface waters make it easier for an El Niño to start, so researchers expect the events to become more frequent. Last year, a model-based study by Wenju Cai, a physical oceanographer with the Commonwealth Scientific and Industrial Research Organisation in Aspendale, Australia, in which Timmermann was involved, suggests that by the end of the century, extreme El Niños such as the 1997–98 event will occur twice as often as they have in recent decades (W. Cai *et al. Nature Clim. Change* **4,** 111–116; 2014).
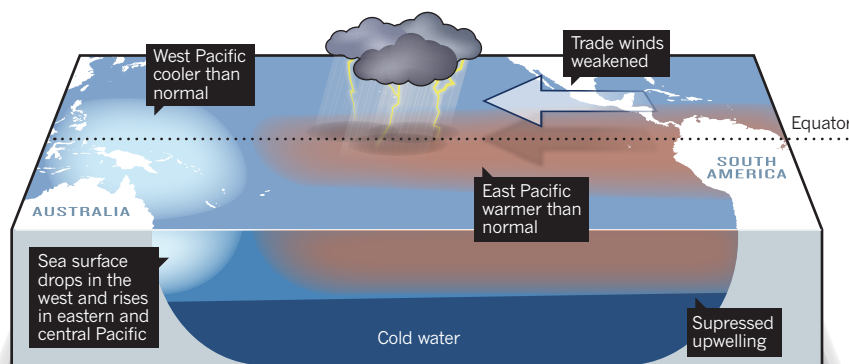
PROBLEM CHILD

The Pacific warming was first described in the late 1880s by a Peruvian Navy captain who reported on an unusually warm '*corriente del Niño*' (ocean current of the Christ Child), so named because it appeared around Christmas time. For a long time, El Niño was thought to be a local phenomenon off Peru and Ecuador. But measurement campaigns during the International Geophysical Year 1957–58, which coincided with a major El Niño, revealed that the phenomenon spans the whole Pacific Ocean. Over the decades since, research on El Niño and La Niña has shown how conditions

community by collecting data on factors including water temperature, salinity and density as it made its servicing runs to the buoys. With the ship no longer available, data collected from buoys and autonomous floats are not sufficient to study the subtle changes in currents and ocean mixing that may be involved in El Niño evolution, says Timmermann.

Researchers have other vessels available, such as the RV *Falkor*, on which Richards made his trip. The RV *Kilo Moana* hosted a second University of Hawaii team in the equatorial Pacific in August and September. Oceanographers Brian Popp and Jeffrey Drazen had an unexpected research opportunity: they had planned to study mercury accumulation in marine organisms in a region with strong equatorial upwelling, but the data they collected during their expedition will allow them to examine

Unruly ocean

El Niño warmings develop when equatorial trade winds weaken, letting warm water and thunderstorm activity spread from the west Pacific Ocean to the east.

West Pacific cooler than normal

Trade winds weakened

Equator

SOUTH AMERICA

East Pacific warmer than normal

AUSTRALIA

Sea surface drops in the west and rises in eastern and central Pacific

Cold water

Supressed upwelling

in the ocean and atmosphere reinforce each other to produce warming and cooling.

Although El Niño/La Niña events can cause powerful changes to weather, science-funding agencies have been reluctant to sponsor expensive research expeditions to study them because they are so hard to forecast. Researchers instead rely to a large extent on data from the Tropical Atmosphere Ocean network of buoys strung across the Pacific, which is jointly operated by NOAA and the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). Temperature and salinity data from the array of 70 or so moored buoys allow researchers to detect unusual ocean warming and track the large waves that push warm water eastwards.

But the array is not without problems. Many buoys have failed in recent years, temporarily leaving scientists with data from just 40% of the network. Thanks to repair work, the system is currently back at 80% capacity. But budget cuts in 2012 forced NOAA to decommission a ship, the RV *Ka'imimoana*, which was used for regular maintenance of the array. Over its 16-year service with NOAA, the ship had also made itself invaluable to the El Niño research

how the effects of a strong El Niño might ripple through the marine food web. And yet, says Cai, this year's El Niño — possibly a once-in-a-generation event — is a missed opportunity with respect to going out and documenting the breadth of physical, chemical and biological changes that might occur in the ocean. "It's a pity we can't have more ships at sea," he says.

Help might be on the way. By 2020, NOAA and JAMSTEC hope to have launched a sustained Tropical Pacific Observing System of buoys and satellite instruments to advance understanding of ocean variability and improve weather and climate prediction.

That will be too late to help with the current El Niño, which is expected to peak late this year or early next. In recent months, it has been keeping pace with the most powerful El Niños on record, and westerly wind outbreaks in early October promised to keep the warming going. As a result, forecasters are warning many parts of the globe to prepare for some wild and crazy weather over the next several months. ∎

Quirin Schiermeier *reports for* Nature *from Munich, Germany.*

THE TIME LAB

WHY DOES MODERN LIFE SEEM SO BUSY?

AN OXFORD CENTRE IS TRYING TO FIND ANSWERS WITH THE WORLD'S BIGGEST COLLECTION OF TIME-USE DIARIES.

BY HELEN PEARSON

In 1961, when more and more people were buying television sets to go with their radios, the BBC wanted to work out the best times to air its programmes. So its audience-research department decided to ask a sample of people across the United Kingdom to record what they were doing every half hour of the day, and to indicate whether the TV or radio was on.

The result was a trove of 2,363 diaries filled with the everyday details of British lives. "8 a.m., Eating breakfast," read one; "8.30 a.m., Taking children to school; 9 a.m., Cleaning away, washing up and listening to Housewives' Choice" — a popular radio record-request programme of the day.

Today, these files are part of the biggest collection of time-use diaries in the world, kept by the Centre for Time Use Research at the University of Oxford, UK. The centre's holdings have been gathered from nearly 30 countries, span more than 50 years and cover some 850,000 person-days in total. They offer the most detailed portrait ever created of when people work, sleep, play and socialize — and of how those patterns have changed over time. "It certainly is unique," says Ignace Glorieux, a sociologist at the Dutch-speaking Free University of Brussels. "It started quite modest, and now it's a huge archive."

The collection is helping to solve a slew of scientific and societal puzzles — not least, a paradox about modern life. There is a widespread perception in Western countries that life today is much busier than it once was, thanks to the unending demands of work, family, chores, smartphones and e-mails. But the diaries tell a different story: "We do not get indicators at all that people are more frantic," says John Robinson, a sociologist who works with time-use diaries at the University of Maryland, College Park. In fact, when paid and unpaid work are totted up, the average number of hours worked every week has not changed much since the 1980s in most countries of the developed world.

Epidemiologists, meanwhile, are mining the diaries to explain how lifestyle changes are contributing to a rise in many chronic diseases. The diaries "were the greatest asset I could possibly have", says physiologist

Edward Archer at the University of Alabama at Birmingham, who used the data in a 2013 study[1] of obesity.

Now, the Oxford centre is testing a major update to its 50-year-old methods. In addition to asking people to complete a handwritten diary, it last year began giving them an electronic fitness tracker and a small camera that snaps a stream of pictures of their day (see 'The gadget guinea pig'). "The idea is to be a bit more adventurous," says Teresa Harms, a sociology research fellow who is leading the project. "Are new technologies better than what we've been doing all these years?"

TIME MANAGEMENT

Ironically for a scientific institute dedicated to time use, the researchers at the Oxford centre are no better at time management than anyone else. If anything, they are worse. One day in July, students were playing a game of croquet on the lawn outside the centre's home: a stone building in the placid grounds of St Hugh's College. But inside, things were more fraught. One flustered postdoc had slept through her alarm and arrived at 10.33 a.m. — more than an hour late for her meeting. The centre's ebullient founder and co-director, sociologist Jonathan Gershuny, cheerily admitted that his own time management is "terrible" — shortly before locking himself out of his office without his keys, ensuring that he would arrive for his next appointment catastrophically late.

None of this seems to have slowed down Gershuny, who can trace the origins of the centre to the 1970s, when he was starting his career at the University of Sussex in Brighton, UK. Gershuny wanted to predict what society and the economy would look like in future decades — but he realized that there was very little empirical evidence showing how people actually spend their time.

Gershuny started to search for surveys in which people had been asked to record their daily activities. Among his first discoveries were the BBC diaries. Another thousand or so journals, recorded in the 1930s, turned up in a mouldering old tea chest at the university. As Gershuny's diary collection grew, it became obvious that the
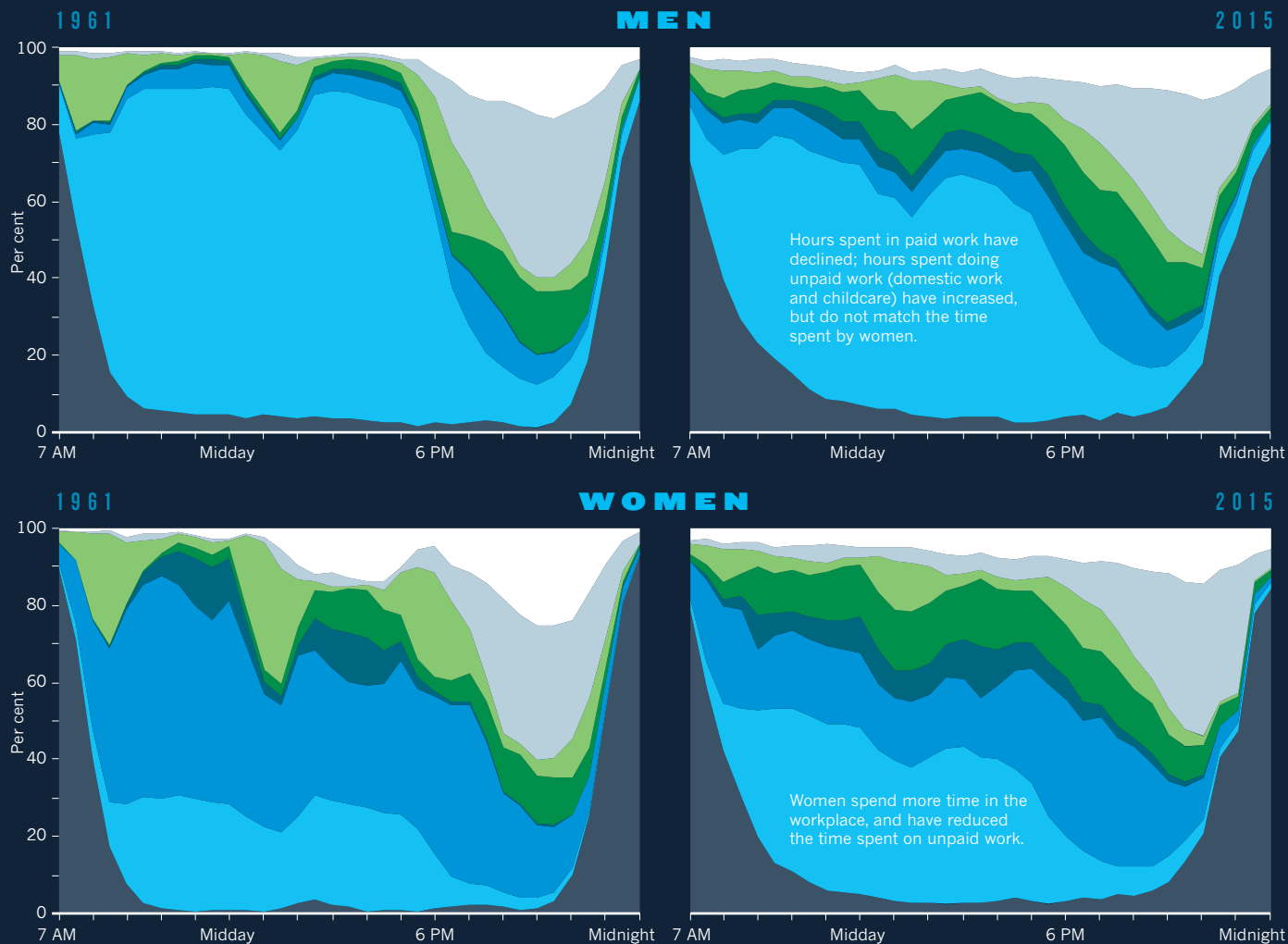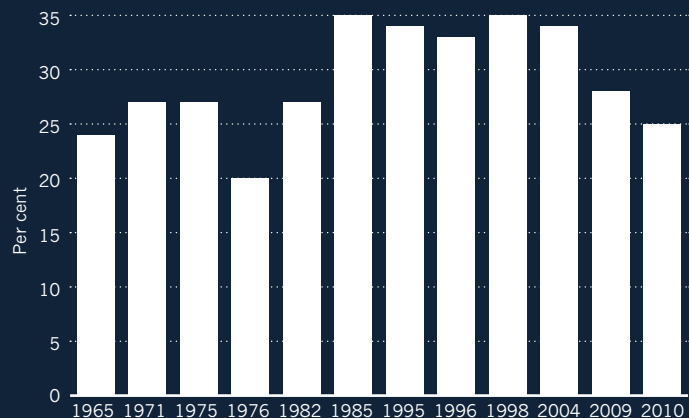
THE TRUTH ABOUT TIME

Analysis of diaries reveals how the average use of time has changed on a standard weekday between 1961 and 2015 in the United Kingdom (similar patterns are seen in other developed countries). In broad terms, the data show a slight growth in leisure time for men and women, and that patterns of paid work are changing for both.

Legend:
- Sleeping or personal
- Work away from home
- Work in or around home
- Travel or shopping
- Leisure away from home
- Eating at home
- TV, radio, video
- Other home leisure

MEN

1961 / **2015**

Per cent — 0, 20, 40, 60, 80, 100

7 AM — Midday — 6 PM — Midnight

Hours spent in paid work have declined; hours spent doing unpaid work (domestic work and childcare) have increased, but do not match the time spent by women.

WOMEN

1961 / **2015**

Per cent — 0, 20, 40, 60, 80, 100

7 AM — Midday — 6 PM — Midnight

Women spend more time in the workplace, and have reduced the time spent on unpaid work.

HOW RUSHED ARE YOU?

Despite a perception that life has become busier, the number of people in a US survey who report feeling 'always rushed' has fallen in the past decade.

Per cent — 0, 5, 10, 15, 20, 25, 30, 35

1965, 1971, 1975, 1976, 1982, 1985, 1995, 1996, 1998, 2004, 2009, 2010

HOW MANY HOURS DO YOU WORK?

A comparison of estimated work hours with actual hours (based on US time-use diaries, 2003–07) showed that people often overestimate how much they work.

Gap between estimated hours and actual hours — −10, 0, 10, 20, 30

People who think that they work the most hours tend also to overestimate the most.

Number of work hours estimated — 1–14, 15–24, 25–34, 35–39, 40, 41–45, 46–54, 55–64, 65–74, 75+
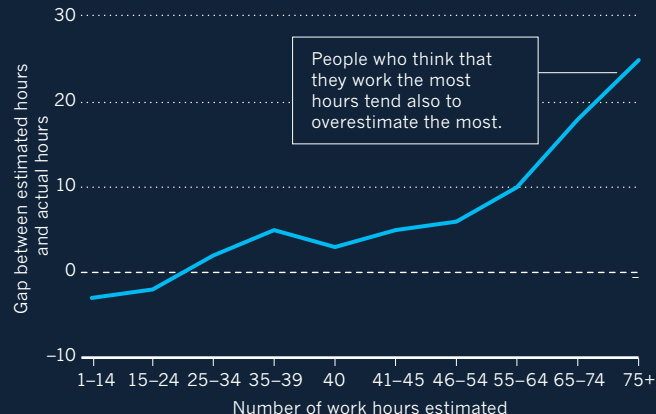
records had been gathered by many different investigators around the world for many different purposes. To align the data and make meaningful comparisons, he would have to put them into a standardized form. So in the 1980s — by then working at the University of Bath, UK — he developed the Multinational Time Use Study: a system in which every activity is given one of 41 codes (gardening, 9; sleeping, 16; relaxing, 36). In an early project that assessed diaries from the United States and the United Kingdom, Gershuny and Robinson showed in 1988 that women in both nations were spending less time on domestic work, whereas men were doing slightly more — a consequence of women's increasing entry into paid employment and of shifting societal norms[2].

By the early 2000s, many countries had started to collect standardized time-use data; the US Bureau of Labor Statistics started gathering them annually in 2003. These efforts were driven by a growing global interest in understanding the impacts of time use on economies and on well-being.

But the diary bank still remained something of a side line for Gershuny until 2008, when he won funding to develop a centre at Oxford dedicated to time-use research. Then, in 2013, the centre was awarded two major grants: €2.5 million (US$2.8 million) from the European Research Council and £3.7 million (US$5.7 million) from the UK Economic and Social Research Council (ESRC), to exploit the

diary bank and to launch a major collection of time-use diaries in the United Kingdom.

"We were paid suddenly to do the things that I'd been agitating to do," Gershuny says. And one of those things has been to find out why some people feel so busy all the time.

NO TIME

In 1930, the economist John Maynard Keynes wrote an essay predicting life 100 years ahead. The United States and Europe would be so prosperous that people would work just 15 hours a week, he said, and the main concern for "our grandchildren" would be how to fill their copious leisure time.

That's not quite how things are turning out — something that Gershuny started to think about in the early 2000s. He was feeling desperately busy — more so than in the past — and people around him were complaining that they were stressed out and working harder as well. Books on the matter were proliferating, with titles such as *Fighting for Time*[3], *Busier than Ever*[4] and *Work Without End*[5]. Survey data hinted at the problem too: in the United States, the proportion of people reporting[6] that they 'always' felt rushed was 24% in 1965 but 34% in 2004.

Yet when researchers used diary data to look into the matter, a

The diary
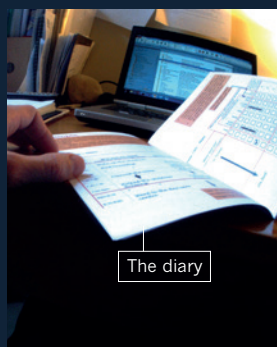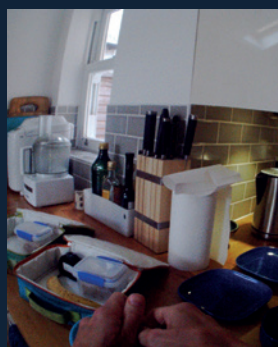
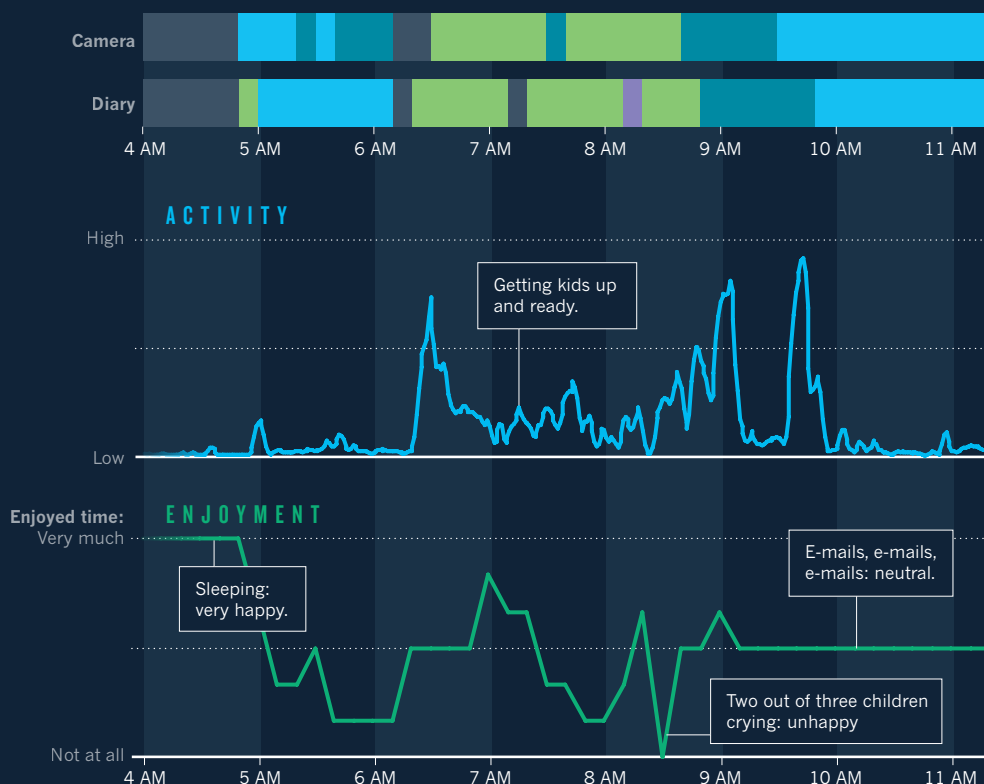5:37 AM   6:49 AM   7:19 AM   9:14 AM   10:56 AM

THE GADGET GUINEA PIG

DIARY OF A NATURE EDITOR

On 16 July 2015, I wore an accelerometer that tracks movement and a camera that took three images per minute. I also recorded what I was doing — and how much I was enjoying it — in a written diary. The Oxford Centre for Time Use Research in the United Kingdom is collecting gadget diaries in this way to find out if they produce more-useful information than do conventional paper diaries. **H.P.**

Camera

Diary

4 AM   5 AM   6 AM   7 AM   8 AM   9 AM   10 AM   11 AM

ACTIVITY

High

Getting kids up and ready.

Low

Enjoyed time:

ENJOYMENT

Very much

Sleeping: very happy.

E-mails, e-mails, e-mails: neutral.

Two out of three children crying: unhappy

Not at all

4 AM   5 AM   6 AM   7 AM   8 AM   9 AM   10 AM   11 AM

different picture emerged. Analyses showed that people in many countries routinely overestimate the amount of time that they spend working — in the United States, by some 5–10% on average[7] (see 'The truth about time'). But those who work longer hours tend to overestimate by the most: people who guess that they work 75-hour weeks, for example, can be over by more than 50%, and those of certain professions — teachers, lawyers, police officers — overestimate by more than 20%. (Scientists were not the worst exaggerators: they estimate working close to 42 hours per week on average, whereas diaries clock them at 39 hours)[8].

In a 2005 study[9], Gershuny compared the BBC diaries from 1961 with UK diaries collected in 1983–84 and 2001, adding up the number of minutes per day spent on paid work, unpaid work (such as chores around the house) and other activities. He wanted to know whether people were actually working longer hours than they did 40 years ago.

The answer was that it depends. Men had reduced the number of hours they spent on paid work, increased those in unpaid work and overall came out ahead, with just under 50 minutes more free time per day. Women were doing more paid work — again reflecting their movement into the workplace over the decades — and less unpaid work, producing little change overall.

Studies in the United States and western European countries have

shown similar patterns: little overall change in work time and, at least in some studies and groups, a slight growth in leisure time. All in all, there is little support for the idea that everyone is working harder than ever before. "When you look at national averages of time-use data, it doesn't really show up," says Oriel Sullivan, a sociologist who now co-directs the centre with Gershuny.

But certain groups have experienced a different trend. According to analyses by Gershuny, Sullivan and other time-use researchers, two demographic groups are, in fact, working harder. One consists of employed, single parents, who put in exceptionally long hours compared to the average; the other comprises well-educated professionals[10], particularly those who also have small children. People in this latter group find themselves pushed to work hard and under societal pressure to spend quality time with their kids. "The combination of those pressures has meant that there is this group for which time pressure is particularly pertinent," Sullivan says.

These findings, the researchers say, could help to explain why there is a widespread perception that life is busier for everyone. Sullivan and Gershuny propose that the time-squeezed professional group includes many of the academics who study and discuss the phenomenon, as well as the journalists who write about it — in other words,

12:28 PM   4:08 PM   5:48 PM   6:52 PM   9:56 PM

The camera

Midday   1 PM   2 PM   3 PM   4 PM   5 PM   6 PM   7 PM   8 PM   9 PM   10 PM   11 PM   Midnight

■ Sleep, personal care   ■ Working   ■ House and family care   ■ Travel and unspecified   ■ Social life   □ Hobbies, computing   ■ Mass media

Walking to and from the station.

Taking the stairs.

At the computer.

Putting toddler to bed: pretty happy.

Dull meeting: unhappy.

Midday   1 PM   2 PM   3 PM   4 PM   5 PM   6 PM   7 PM   8 PM   9 PM   10 PM   11 PM   Midnight
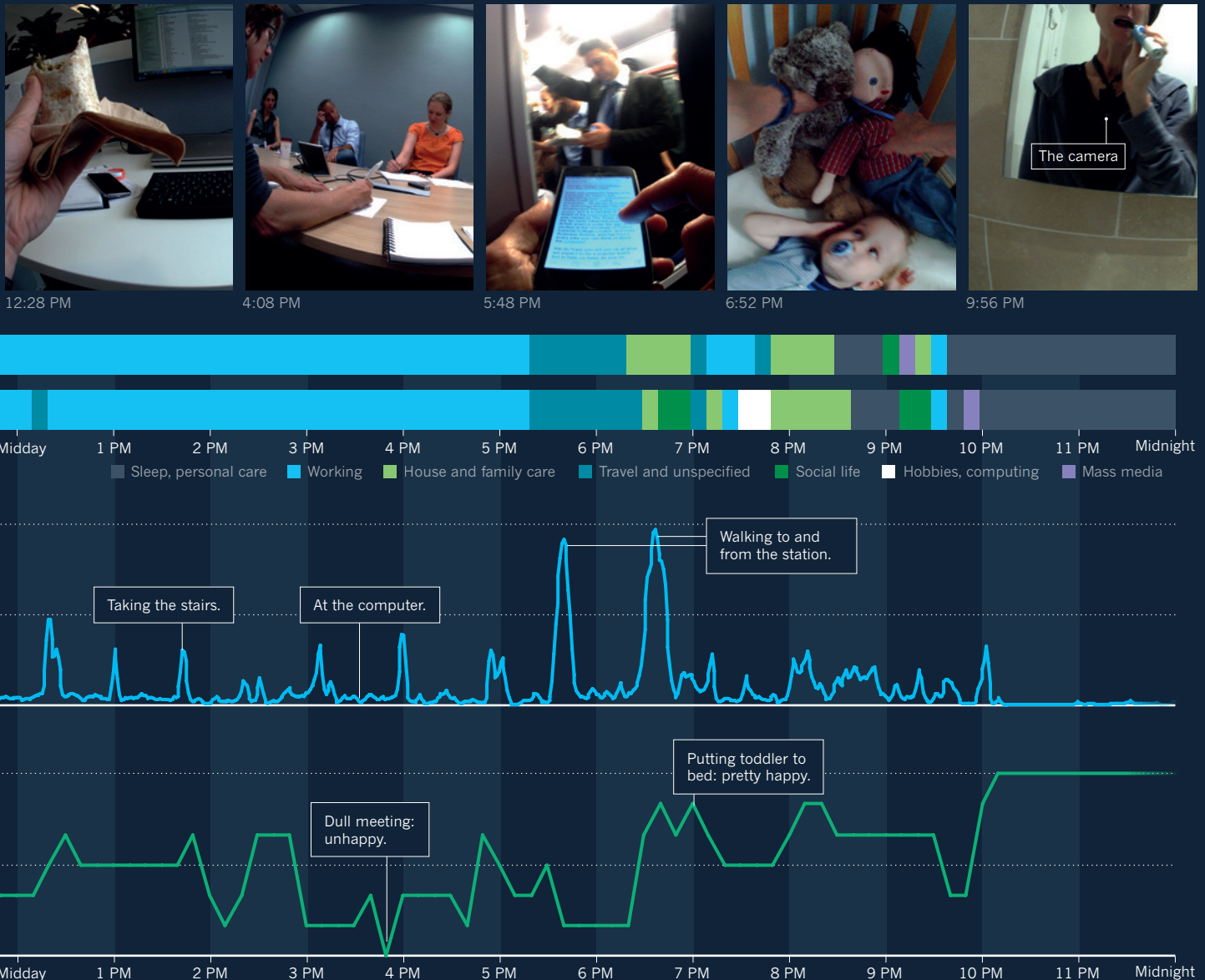
the people in society with a loud voice.

But Gershuny suggests that changing attitudes to work and leisure may also play a part. In nineteenth-century Europe, having ample leisure time signified a person of high social status: one philosopher described the literary types in Paris around 1840, who had such an abundance of time that it was fashionable to walk a turtle on a leash through the arcades.

In the twenty-first century, the situation has reversed, so that being busy is a signal of a privileged social position — and therefore an impression that some people are keen to give. Gershuny calls being busy in the modern day a "badge of honour", and Glorieux agrees. "The first thing we say when we meet people is 'I'm busy'," he says. "Suppose we say, 'I'm not busy; I have nothing to do; I was watching some TV.' It's not what people want to say."

People might also feel busier because of an increase in multitasking, especially with computers and smartphones. The US time-use diaries are poor at recording how long people engage with their devices, says Robinson — in part, he suspects, because they have become so prevalent that people don't even report when they are on them.

FAMILY TIME

The diary bank at Oxford has revealed other changes in how people use their time. In 2011, Oxford centre postdoc Evrim Altintaş drew on diaries collected in the United States between 1965 and 2013 to examine how much time parents spend on 'developmental childcare' — engaging with children through reading, talking and helping with homework[11]. Activities of this type are strongly associated with better educational scores, behaviour and other positive outcomes in later life.

Altintaş found that parents overall were spending more time on developmental childcare in the 2000s than they were in the 1960s and 1970s, but she found a bigger increase among parents who both had university degrees than for those who had high-school diplomas or less. She estimated that a child born to a highly educated mother in the early 2000s would receive 27 minutes more developmental childcare per day than one born to less-educated parents — adding up to 657 extra hours of focused attention for that child during the first four years of life. "That puts the children who are born to less-educated parents at a real disadvantage," Altintaş says.

The diaries have also exposed trends that could affect the health of adults. In his study on obesity[1], Archer analysed more than 50,000 diary days collected between 1965 and 2010 and divided women's time into paid work, household work, personal care and free time. Then he calculated what that meant for the amount of energy they were burning up. The results showed that women in 2010 were spending around 12 hours less per week on cooking, cleaning, laundry and other domestic work

than women in 1965, and that had shifted towards more-sedentary pursuits such as using a computer. As a result, the team estimated that working women today are burning some 130 kilocalories per day less than those in the 1960s, and they proposed that this could be one explanation for the rise of obesity in the United States. (Archer stresses that he is not saying that women should do more housework; rather, the work reinforces public-health advice encouraging more physical activity of any kind.)

> ## "THE FIRST THING WE SAY WHEN WE MEET PEOPLE IS 'I'M BUSY.'"

At Oxford, Gershuny and Harms are attempting to carry out more-detailed analyses of energy use in collaboration with researchers at the US National Cancer Institute in Rockville, Maryland. Harms has been matching up entries in a selection of diaries to a list, widely used in research, of more than 800 activities alongside estimates of the energy burned doing each. (It includes entries as specific as playing darts, coalmining, whirlpool-sitting and casino gambling.) The study, which is still under way, has so far shown that a gym session or other structured work-out accounts for only a small fraction of the energy that a person typically burns each day. Activities such as paid work and childcare often burn more — because even if they are less physically intense, they take up longer periods of time. "The real metabolic activity is built up during the working day," Harms says.

NEXT-GENERATION DIARIES

Since Gershuny started his diary bank, time-use research has become a thriving industry: there are now several hundred researchers in the field. But time-use diaries have weaknesses, and the biggest is that they can be wrong: people quickly forget what they were doing and record their days inaccurately — or they can lie. The desire to improve accuracy is one of the motivations behind CAPTURE-24: Gershuny and Harms's project to collect a new generation of diaries using some of the latest gadgets around.

So far, about 150 people have each spent 24 hours wearing a watch-like accelerometer

strapped to their wrists and a small camera, which takes around three pictures per minute, hung around their necks. They also jot down what they are doing for every 10-minute slot of the day in a conventional paper diary. The aim is to see whether the devices can provide more-useful information for the researchers than a standard paper diary alone.

The accelerometer should collect more-accurate data on body movement and energy use, one reason that the project has earned support from biomedical-research funders the British Heart Foundation and the Wellcome Trust, both in London, as well as the ESRC. The photos could keep a more faithful record of when and what people eat — food diaries are notoriously unreliable — or reveal important nuances in people's interactions with children. (It is harder to say that you were focusing on childcare when the camera pictures show that you were checking your phone.)

In her preliminary analyses, Harms has found that gadget diaries and paper diaries show the same sequence of events, but that the gadgets reveal details that paper diaries missed. Most researchers in the field agree that the future lies in collecting data through phones and other devices. "Maybe this will bring a new boost to time-use research," Glorieux says. He anticipates a situation in which reams of diary data — such as location, heart rate, calories burned and even ambient noise — are collected through phones and linked-up gadgets.

The researchers at Oxford are keen to grow their diary collection in other ways. They recently added ones from China, South Korea and India, and they are trying to include more from Eastern European and developing countries. And Gershuny holds out hope that there are more tea chests of old diaries still waiting to be found. Then, scientists can begin to examine cultural differences in how people from different regions work, rest and play.

So many questions, so much data, so little time. Clearly, doing all this is going to take a lot more than Keynes's 15 hours a week. But the scientists hope to get there, by taking it one day at a time. ■

Helen Pearson is Chief Features Editor for Nature.

1. Archer, E. *et al. PLoS ONE* **8**, e56620 (2013).
2. Gershuny, J., Robinson, J. P. *Demography* **25**, 537–552 (1988).
3. Epstein, C. & Kalleberg, A. (eds) *Fighting for Time* (Russell Sage Foundation, 2014).
4. Darrah, C., Freeman, J. & English-Lueck, J. *Busier than Ever* (Stanford Univ. Press, 2007).
5. Hunnicutt, B. *Work without End* (Temple Univ. Press, 1998).
6. Robinson, J. P. *Soc. Indic. Res.* **113**, 1091–1104 (2013).
7. Robinson, J. P., Martin, S., Glorieux, I. & Minnen, J. *Mon. Labor Rev.* **134**, 43–53 (2011).
8. Robinson, J. P. & Gershuny, J. *Occupational Differences in Estimates of Time at Work* (Maryland Population Research Center, 2011).
9. Gershuny, J. *Soc. Res.* **72**, 287–314 (2005).
10. Gershuny, J. *Soc. Indic. Res.* **93**, 37–45 (2009).
11. Altintaş, E. *J. Marriage Fam.* (in the press).

COMMENT

China's cities are among the world's worst in terms of air quality.

China's choking cocktail

Cleaning up city and indoor air will require a deeper understanding of the unprecedented chemical reactions between pollutants, says **Markku Kulmala**.

Dirty air threatens the health of billions of city dwellers around the world. China's megacities are among the worst, with concentrations of airborne pollutants 10–100 times higher than those in Europe or North America, and occasionally even 1,000 times higher. An estimated 2.5 million people in China die each year from the health effects of indoor and outdoor air pollution[1,2].

Efforts to improve air quality are targeting only the tip of the iceberg. Cities such as Beijing routinely measure levels of particulate matter measuring 10 micrometres (PM_{10}) and 2.5 micrometres ($PM_{2.5}$) in size, as well as a few gases such as sulfur dioxide (SO_2), nitrogen oxides (NO_x), carbon monoxide (CO) and ozone. But urban air is a complex cocktail of chemicals whose poorly understood interactions and feedbacks

may exacerbate health problems. Efforts to reduce one pollutant can have perverse effects on others as conditions change.

The chemistry of China's polluted urban air is unprecedented. Higher populations, heavier industries and modern goods manufacturing, as well as the climatic conditions, make Beijing's smogs markedly different from the 'pea soupers' that afflicted London and other European cities ▶

▶ 50 to 100 years ago. Many atmospheric processes are nonlinear — meaning that the relationship between cause and effect is not proportional. So we do not know and cannot predict which harmful compounds are being formed. Whenever my colleagues and I make measurements with new instruments in China, we find unexpected results. Indoor air quality is equally affected.

The nation's urban air can begin to be cleaned only if a comprehensive monitoring and modelling approach takes atmospheric chemistry into account. To guide decisions, we need to know which hazardous pollutants are present and how they interact to generate secondary pollution.

Here I outline a road map for bringing China's cities up to European air-quality standards within a decade. Actions towards cleaner air, as well as improving health (see 'Deep clean'), will reduce greenhouse-gas and black-carbon (soot) concentrations and enhance freshwater quality[3] and food supply.

TOXIC MIX

China's air pollution has worsened as emissions from industry, energy production and traffic have grown. China is responsible for 30–35% of the global SO_2, NO_x, CO and particulate emissions, and 40% of global particle numbers (PN) in the 20–1,000-nanometre size range (see go.nature.com/uw3jx6). The nation's share of global greenhouse-gas emissions is 29% for carbon dioxide and almost 20% for methane.

Government efforts are under way to reduce the emissions of all these.

But it is impossible to reduce secondary pollutants such as ozone and organic aerosols without a deep understanding of the chains of chemical reactions and physical processes that pollutants undergo. The formation and decay of secondary pollutants depend on temperature, humidity and wind speed as well as other chemicals and particles in the urban atmosphere. There is so much that we do not know. Results are hard to predict because many physical and chemical processes — such as surface chemistry, oxidation, clustering and dynamical effects — happen simultaneously[4].

Attempts to control one pollutant can increase the concentration of others. Measurements in Nanjing, for example, show[5] that reducing NO_x emission could cause a tenfold increase in summer ozone concentrations. Reducing smog increases sunshine levels and temperatures, and alters rain and snowfall patterns[6].

Surprising reactions are going on above Chinese cities that do not occur in cleaner air. For example, small atmospheric

> *"As outdoor air gets cleaner, indoor air quality might even worsen."*

Pollutant levels inside Chinese homes can be hundreds of times higher than those in European homes.

molecular clusters (measuring 1–3 nanometres) are tens of times more concentrated in Shanghai, Nanjing and Beijing than in European cities. Secondary aerosols (containing sulfates, nitrogen compounds and organics) form more readily in Shanghai and Nanjing than existing models predicts[7]. Unknown chemical pathways and physical processes must be occurring[8] that could create new types of oxidant or change the surface properties of aerosols, limiting their ability to take up condensable vapours.

Indoor air quality is also a problem. City dwellers spend more than 90% of their time indoors, particularly if outdoor air quality is poor. Cooking, smoking, heating and furnishings release PM, PN, CO, volatile organic compounds (VOCs) and NO_x,

adding to pollutants drawn in from outside. Little is known about the pollution levels in Chinese homes, but my team's preliminary findings indicate that concentrations of some pollutants are higher than those outdoors (especially for PN and VOCs) and 100–1,000 times higher than inside European homes. My colleagues and I estimate that indoor air pollution contributes almost as much as outdoor smog (which causes more than 1.3 million deaths per year) to the pollution-related death toll.

Again, little is known about the secondary production of chemicals indoors. Ventilation systems filter large particles but not gases such as VOCs, NO_x and SO_2, which can go on to form ultrafine sulfates, nitrates and organic particles. The risk of secondary aerosol production from these gases rises in filtered air because there are fewer grains on which to condense. As outdoor air gets cleaner, indoor air quality might even worsen.
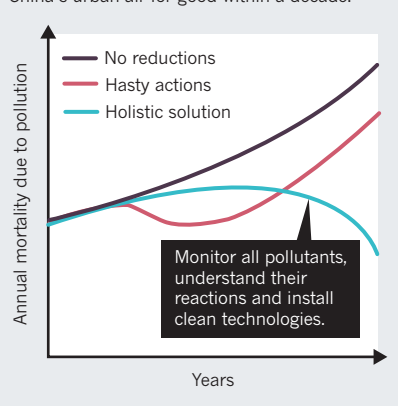
CLEARING THE AIR

It is important to continue efforts to cut pollution. But meeting the Chinese central government's goal of improving urban air quality to levels typical of the United States and Europe requires more: simultaneous tracking of all air pollutants relevant to health and their feedbacks and interactions — over at least a decade. We need to understand how the mixture and its toxicity changes as air quality measures are implemented. Short campaigns are not sufficient.

Central and regional governments, research institutes and universities should collaborate to do the following, with help

DEEP CLEAN

Deaths from air pollution will rise as cities and industries expand. Limiting single pollutants will buy time, but only a concerted effort will clean China's urban air for good within a decade.

Annual mortality due to pollution

— No reductions
— Hasty actions
— Holistic solution

Monitor all pollutants, understand their reactions and install clean technologies.

Years

from atmospheric chemists from around the world.

First, establish and fund a network of 'flagship' stations[9] to monitor: concentrations, fluxes, interactions and feedbacks as well as more general air quality and meteorology data. Around 5–8 such stations (costing between US$7 million and $11 million each) would suffice for a major city. These should be complemented by mobile measurement platforms on cars and aeroplanes, remote sensing of air columns from the ground, satellite observations and smog chambers. Major sources of pollutants can be identified using historic data.

Second, indoor air-quality measurements and monitoring must be conducted concurrently in a representative selection of residential and office buildings.

Third, atmospheric chemists must model secondary-pollutant production pathways and feedback mechanisms under high concentrations of various pollutants. These models must then be compared with observations.

Fourth, the links between air pollutants and mortality and other health effects need to be established. That way the most health-relevant pollutants and their sources can be identified and mitigated first. A database should be developed to track health impacts.

Fifth, long-term sustainable engineering solutions such as improving processes and material flows in industry must be implemented to maintain low levels of air pollution. This will require capacity building across the Chinese authorities and institutes on using air-quality assessment data in decision-making, in developing legislative tools and in clean-air action plans.

Only by understanding atmospheric chemistry will China clean its air. ∎

Markku Kulmala *is professor of aerosol physics at the University of Helsinki, Finland.*
e-mail: markku.kulmala@helsinki.fi

1. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. *Nature* **525**, 367–371 (2015).
2. Lim, S. S. *et al. Lancet* **380**, 2224–2260 (2012).
3. Zhang, J. *et al. Lancet* **375**, 1110–1019 (2010).
4. Guo, S. *et al. Proc. Natl Acad. Sci. USA* **111**, 17373–17378 (2014).
5. Ding, A. J. *et al. Atmos. Chem. Phys.* **13**, 5813–5830 (2013).
6. Ding, A. J. *et al. Atmos. Chem. Phys.* **13**, 10545–10554 (2013).
7. Xiao, S. *et al. Atmos. Chem. Phys.* **15**, 1769–1781 (2015).
8. Kulmala, M. *et al. Annu. Rev. Phys. Chem.* **65**, 21–37 (2014).
9. Hari, P. *et al. Atmos. Chem. Phys. Discuss.* **15**, 21063–21093 (2015).

The mysterious Indus unicorn on a roughly 4,000-year-old sealstone, found at the Mohenjo-daro site.

Cracking the Indus script

Andrew Robinson reflects on the most tantalizing of all the undeciphered scripts — that used in the civilization of the Indus valley in the third millennium BC.

The Indus civilization flourished for half a millennium from about 2600 BC to 1900 BC. Then it mysteriously declined and vanished from view. It remained invisible for almost 4,000 years until its ruins were discovered by accident in the 1920s by British and Indian archaeologists. Following almost a century of excavation, it is today regarded as a civilization worthy of comparison with those of ancient Egypt and Mesopotamia, as the beginning of Indian civilization and possibly as the origin of Hinduism.

More than a thousand Indus settlements covered at least 800,000 square kilometres of what is now Pakistan and northwestern India. It was the most extensive urban culture of its period, with a population of perhaps 1 million and a vigorous maritime export trade to the Gulf and cities such as Ur in Mesopotamia, where objects inscribed with Indus signs have been discovered. Astonishingly, the culture has left no archaeological evidence of armies or warfare.

Most Indus settlements were villages; some were towns, and at least five were substantial cities (see 'Where unicorns roamed'). The two largest, Mohenjo-daro — a World Heritage Site listed by the United Nations — located near the Indus river, and Harappa, by one of the tributaries, boasted street planning and house drainage worthy of the twentieth century AD. They hosted the world's first known toilets, along with ▶

▶ complex stone weights, elaborately drilled gemstone necklaces and exquisitely carved seal stones featuring one of the world's stubbornly undeciphered scripts.

FOLLOW THE SCRIPT

The Indus script is made up of partially pictographic signs and human and animal motifs including a puzzling 'unicorn'. These are inscribed on miniature steatite (soapstone) seal stones, terracotta tablets and occasionally on metal. The designs are "little masterpieces of controlled realism, with a monumental strength in one sense out of all proportion to their size and in another entirely related to it", wrote the best-known excavator of the Indus civilization, Mortimer Wheeler, in 1968[1].

Once seen, the seal stones are never forgotten. I became smitten in the late 1980s when tasked to research the Indus script by a leading documentary producer. He hoped to entice the world's code-crackers with a substantial public prize. In the end, neither competition nor documentary got off the ground. But for me, important seeds were sown.

More than 100 attempts at decipherment have been published by professional scholars and others since the 1920s. Now — as a result of increased collaboration between archaeologists, linguists and experts in the digital humanities — it looks possible that the Indus script may yield some of its secrets.

Since the discovery of the Rosetta Stone in Egypt in 1799, and the consequent decipherment of the Egyptian hieroglyphs beginning in the 1820s, epigraphers have learnt how to read an encouraging number of once-enigmatic ancient scripts. For example, the Brahmi script from India was 'cracked' in the 1830s; cuneiform scripts (characterized by wedge-shaped impressions in clay) from Mesopotamia in the second half of the nineteenth century; the Linear B script from Greece in the 1950s; and the Mayan glyphs from Central America in the late twentieth century.

"No firm information is available about its underlying language."

Several important scripts still have scholars scratching their heads: for example, Linear A, Etruscan from Italy, Rongorongo from Easter Island, the signs on the Phaistos Disc from the Greek island of Crete and, of course, the Indus script.

In 1932, Flinders Petrie — the most celebrated Egyptologist of his day — proposed an Indus decipherment on the basis of the supposed similarity of its pictographic principles to those of Egyptian hieroglyphs. In 1983, Indus excavator Walter Fairservis at the American Museum of Natural History in New York City, claimed in *Scientific American*[2] that he could read the signs in a form of ancient Dravidian: the language

Mohenjo-daro existed at the same time as the civilizations of ancient Egypt, Mesopotamia and Crete.

family from southern India that includes Tamil. In 1987, Assyriologist James Kinnier Wilson at the University of Cambridge, UK, published an 'Indo-Sumerian' decipherment, based on a comparison of the Indus signs with similar-looking ones in cuneiform accounting tablets from Mesopotamia.

THREE PROBLEMS

In the 1990s and after, many Indian authors — including some academics — have claimed that the Indus script can be read in a form of early Sanskrit, the ancestral language of most north Indian languages including Hindi. In doing so, they support the controversial views of India's Hindu nationalist politicians that there has been a continuous, Sanskrit-speaking, Indian identity since the third millennium BC.

Whatever their differences, all Indus researchers agree that there is no consensus on the meaning of the script. There are three main problems. First, no firm information is available about its underlying language. Was this an ancestor of Sanskrit or Dravidian, or of some other Indian language family, such as Munda, or was it a language that has disappeared? Linear B was deciphered because the tablets turned out to be in an archaic form of Greek; Mayan glyphs because Mayan languages are still spoken. Second, no names of Indus rulers or personages are known from myths or historical records: no equivalents of Rameses or

Ptolemy, who were known to hieroglyphic decipherers from records of ancient Egypt available in Greek.

Third, there is, as yet, no Indus bilingual inscription comparable to the Rosetta Stone (written in Egyptian and Greek). It is conceivable that such a treasure may exist in Mesopotamia, given its trade links with the Indus civilization. The Mayan decipherment started in 1876 using a sixteenth-century Spanish manuscript that recorded a discussion in colonial Yucatan between a Spanish priest and a Yucatec Mayan-speaking elder about ancient Mayan writing.
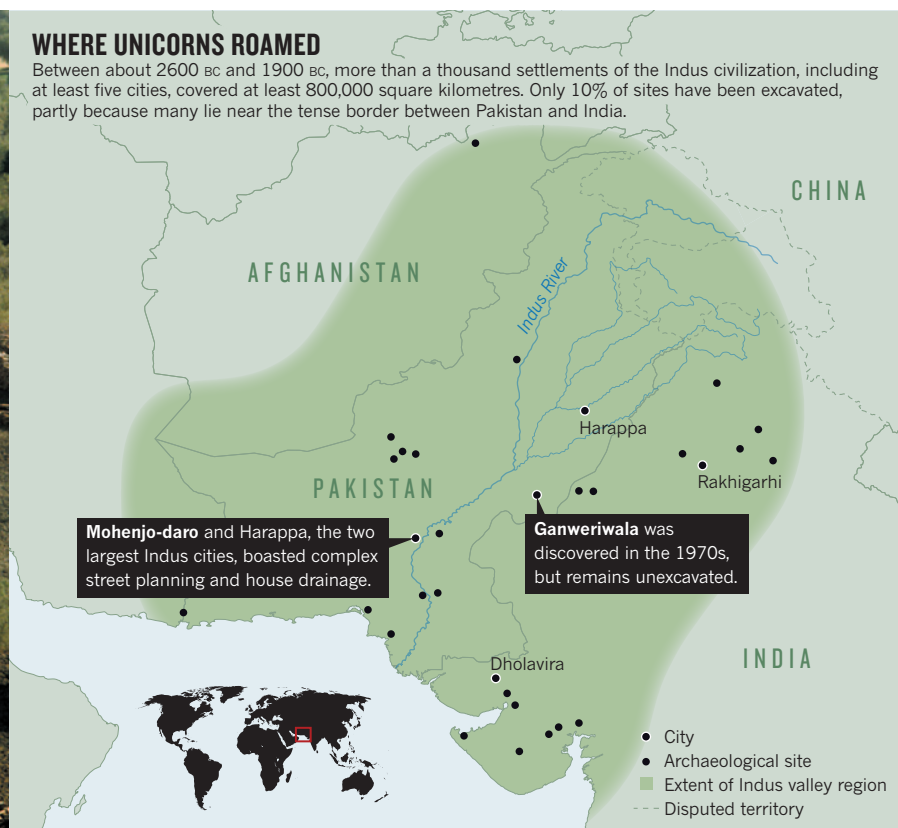
WHAT WE KNOW

Indus scholars have achieved much in recent decades. A superb three-volume photographic corpus[3] of Indus inscriptions, edited by the indefatigable Asko Parpola, an Indologist at the University of Helsinki, was published between 1987 and 2010 with the support of the United Nations Educational, Scientific and Cultural Organization; a fourth and final volume is still to come. The direction of writing — chiefly right to left — has been established by analysis of the positioning of groups of characters in many differing inscriptions. The segmentation of texts containing repeated sequences of characters, syntactic structures, the numeral system and the measuring system are partly understood.

Views vary on how many signs there are

WHERE UNICORNS ROAMED
Between about 2600 BC and 1900 BC, more than a thousand settlements of the Indus civilization, including at least five cities, covered at least 800,000 square kilometres. Only 10% of sites have been excavated, partly because many lie near the tense border between Pakistan and India.

CHINA

AFGHANISTAN

Indus River

PAKISTAN

Harappa

Rakhigarhi

Mohenjo-daro and Harappa, the two largest Indus cities, boasted complex street planning and house drainage.

Ganweriwala was discovered in the 1970s, but remains unexcavated.

INDIA

Dholavira

• City
• Archaeological site
  Extent of Indus valley region
- - - Disputed territory

in the Indus script. In 1982, archaeologist Shikaripura Ranganatha Rao published a Sanskrit-based decipherment with just 62 signs[4]. Parpola put[5] the number at about 425 in 1994 — an estimate supported by the leading Indus script researcher in India, Iravatham Mahadevan. At the other extreme is an implausibly high estimate[6] of 958 signs, published this year by Bryan Wells, arising from his PhD at Harvard University in Cambridge, Massachusetts.

Nevertheless, almost every researcher accepts that the script contains too many signs to be either an alphabet or a syllabary (in which signs represent syllables), like Linear B. It is probably a logo-syllabic script — such as Sumerian cuneiform or Mayan glyphs — that is, a mixture of hundreds of logographic signs representing words and concepts, such as &, £ and %, and a much smaller subset representing syllables.

As for the language, the balance of evidence favours a proto-Dravidian language, not Sanskrit. Many scholars have proposed plausible Dravidian meanings for a few groups of characters based on Old Tamil, although none of these 'translations' has gained universal acceptance.

A minority of researchers query whether the Indus script was capable of expressing a spoken language, mainly because of the brevity of inscriptions. The carvings average five characters per text, and the longest has only 26. In 2004, historian Steve Farmer, computational linguist Richard

Sproat (now a research scientist at Google) and Sanskrit researcher Michael Witzel at Harvard University caused a stir with a joint paper[7] comparing the Indus script with a system of non-phonetic symbols akin to those of medieval European heraldry or the Neolithic Vinča culture from central and southeastern Europe[8].

This theory seems unlikely, for various reasons. Notably, sequential ordering and an agreed direction of writing are universal features of writing systems. Such rules are not crucial in symbolic systems. Moreover, the Indus civilization must have been well aware through its trade links of how cuneiform functioned as a full writing system.

Nevertheless, the brevity of Indus texts may indeed suggest that it represented only limited aspects of an Indus language. This is true of the earliest, proto-cuneiform, writing on clay tablets from Mesopotamia, around 3300 BC, where the symbols record only calculations with various products (such as barley) and the names of officials.

DIGITAL APPROACH
The dissident paper has stimulated some fresh approaches. Wells — a vehement believer that the Indus script is a full writing system — working with the geoinformation scientist Andreas Fuls at the Technical University of Berlin, has created the first, publicly available, electronic corpus of Indus texts (see www.archaeoastronomie.de). Although not complete, it includes all the texts

from the US-led Harappa Archaeological Research Project.

A group led by computer scientist Rajesh Rao at the University of Washington in Seattle has demonstrated the potential of a digital approach. The team has calculated the conditional entropies — that is, the amount of randomness in the choice of a token (character or word) given a preceding token — in natural-language scripts, such as Sumerian cuneiform and the English alphabet, and in non-linguistic systems, such as the computer programming language Fortran and human DNA. The conditional entropies of the Indus script seem to be most similar to those of Sumerian cuneiform. "Our results increase the probability that the script represents language," the Rao group has written[9]. Sproat strongly disagrees[10].

On the ground in Pakistan and India, more inscriptions continue to be discovered — although not, as yet, any texts longer than 26 characters. Unfortunately, less than 10% of the known Indus sites have been excavated. The difficulty — apart from funding — is the politically troubled nature of the region. Many of the most promising unexcavated sites lie in the Pakistani desert region of Cholistan near the tense border with India. One such is the city of Ganweriwala, discovered in the 1970s and apparently comparable in size with Mohenjo-daro and Harappa.

If these sites, and some others within Pakistan and India, were to be excavated, there seems a reasonable prospect of a widely accepted, if incomplete, decipherment of the Indus script. It took more than a century to decipher the less challenging Mayan script, following several false starts, hiatuses and extensive excavation throughout the twentieth century. Indus-script decipherers have been on the much barer trail — older by two millennia — for less than a century, and excavation of Indus sites in Pakistan has stagnated in recent decades. ∎

Andrew Robinson *is a science writer based in London. He is the author of* Lost Languages: The Enigma of the World's Undeciphered Scripts *and, most recently,* The Indus: Lost Civilizations.
e-mail: andrew.robinson33@virgin.net

1. Wheeler, M. *The Indus Civilization* 3rd edn 101 (Cambridge Univ. Press, 1968).
2. Fairservis, W. A. *Sci. Am.* **248,** 58–66 (1983).
3. Parpola, A. *et al.* (eds) *Corpus of Indus Seals and Inscriptions* Vols 1–3.1 (Suomalainen Tiedeakatemia, 1987, 1991, 2010).
4. Rao, S. R. *The Decipherment of the Indus Script* (Asia Publishing, 1982).
5. Parpola, A. *Deciphering the Indus Script* (Cambridge Univ. Press, 1994).
6. Wells, B. K. *The Archaeology and Epigraphy of Indus Writing* (Archaeopress, 2015).
7. Farmer, S., Sproat. R & Witzel, M. *Electron. J. Vedic Stud.* **11,** 19–57 (2004).
8. Lawler, A. *Science* **306,** 2026–2029 (2004).
9. Rao, R. P. N. *et al. Science* **324,** 1165 (2009).
10. Sproat, R. *Language* **90,** 457–481 (2014).

A health nurse in Ghana washes her hands before examining a baby.

BACTERIOLOGY

Pathogens in perspective

Andrew Jermy travels with Hugh Pennington on the arc of humanity's long, troubled relationship with microorganisms.

Before opening Hugh Pennington's *Have Bacteria Won?*, readers of newspaper headlines might presume that his answer is 'yes'. But for the most part, the eminent bacteriologist comes to the opposite conclusion in this thought-provoking study that documents the history of human interactions with infectious disease and how current fears of impending doom have developed.

Pennington draws on personal experience and illuminating case studies — such as the United Kingdom's experience with bovine spongiform encephalopathy (BSE) and variant Creutzfeldt–Jakob disease — to show how the public perception and clinical reality of infectious disease can be at odds. He upbraids researchers, journalists and editors (such as me) for using the hyperbolic language of war (fight, struggle, arms race) to describe our relationship with the microorganisms that colonize and infect our bodies. Such language injects drama and elevates the importance of events, much more than dry but accurate descriptions of the consequences of interactions between microbe, host, immune response and treatment.

As a result, the fear of microbial life in the collective mind is often vastly out of proportion to the risk. As Pennington puts it, "the media behave like a cheap refracting telescope, focusing on an object of interest but magnifying it with a good deal of aberration and fuzziness at the edges". Witness the media hysteria in 2014 when US nurse Kaci Hickox returned to Maine from Sierra Leone after working with Médecins Sans Frontières (also known as Doctors Without Borders). Hickox was wrongly suspected of infection with Ebola, and her return set in train legal proceedings relating to her quarantine.

Oddly, Pennington then fails to heed his own critique about rhetoric. He tours some of our "victories" (against smallpox, diphtheria and syphilis); the "advance" of microorganisms such as *Escherichia coli* and MRSA — methicillin-resistant *Staphylococcus aureus* — through horizontal transfer of

Have Bacteria Won?
HUGH PENNINGTON
Polity: 2015

toxin-encoding genes or selection for antibiotic resistance; and the "battles" in which human actions have helped microorganisms (including *Salmonella* and the organisms that cause anthrax and legionnaires' disease). He deftly weaves historical vignettes into the greater journey. These include early efforts to control smallpox in the eighteenth century, led by Lady Mary Wortley Montagu, US minister Cotton Mather and latterly Edward Jenner; the benefits of improved water availability (originally intended to support trade and fight fire); sanitation, diet and pasteurization in the nineteenth century; and on to the discovery of antibiotics in the twentieth century, right up to outbreaks of carbapenem-resistant Enterobacteriaceae and severe acute respiratory syndrome (SARS) coronavirus in the modern era. The arc of that story and Pennington's accessible prose grip throughout.

Pennington points out that infection with antibiotic-resistant bacteria is not new. It has followed closely in the footsteps of all antibiotics since penicillin — discovered by Alexander Fleming in 1928 — was developed as a treatment by Howard Florey and Ernst Chain. In another bout of debunking,

Pennington argues that predictions of a coming antibiotics Armageddon leading to a substantial increase in infection-related deaths are greatly exaggerated. On this point, I take a more cautious line. It is true that careful management and aseptic technique can have an important role in husbanding a dwindling supply of drugs effective against the most serious infections. However, Pennington does not devote sufficient space to the factors that have led the antibiotic-development pipeline to dry up in recent years.

He almost trivializes the difficulties in identifying relevant natural products or chemical constructs and developing them into usable drugs, simply writing: "New antimicrobials will be very welcome. Getting them ready for rollout will be expensive and will take years." And he skates over structural problems in the pharmaceutical industry: we urge pharma to develop antimicrobials while simultaneously planning to limit their use drastically. In the twentieth century, drugs came along in time to take over when resistance arose; whether that will be the future pattern is uncertain.

Pennington also skimps on coverage of microscopic eukaryotic pathogens, such as the malarial parasite *Plasmodium falciparum* or the fungi that cause cryptococcal meningitis. His only mention of malaria is in relation to Nobel-prizewinning Austrian physician Julius Wagner-Jauregg's use of *Plasmodium* infection as an experimental antimicrobial agent to trigger the inflammation necessary to kill *Treponema*, the spirochaete that causes syphilis. Yet malaria currently kills more than 500,000 people a year, and the spread in southeast Asia of resistance to the only effective antimalarials is of global concern.

As Pennington admits, *Have Bacteria Won?* is intentionally biased by his personal experience as an infectious-disease specialist working in the United Kingdom. It would be unreasonable to expect comprehensive coverage in an overview for the generalist. But he could have better explored the idea that developed countries are over-fearful about infectious diseases, whereas developing nations — struggling with poor sanitation and inadequate clean water, nutrition and health care — are at greater, and globally significant, risk.

The book's title notwithstanding, Pennington extends his analyses to diseases caused by viruses, prions and eukaryotic parasites. Microbiologists grind their teeth when a well-intentioned news report refers to a bacterial infection as caused by a virus, or vice versa, so why sow more confusion? However, these few concerns do not detract from what is an entertaining and very well-written primer on the human–microbe relationship — one of the oldest pairings on Earth. ∎

Andrew Jermy *is chief editor of* Nature Microbiology.
e-mail: a.jermy@nature.com

Books in brief
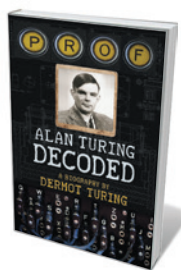
Prof: Alan Turing Decoded
Dermot Turing THE HISTORY PRESS *(2015)*
Computing pioneer Alan Turing has been justly and amply lauded, not least in these pages (see *Nature* **515,** 195–196 (2014) and nature.com/turing). Now, a biography by his nephew, Dermot Turing, offers new sources and a refreshingly familial tone. We see the stubbornly original young Alan finding his way in maths and society; mentors such as Max Newman, who kick-started Turing's obsession with machines; the design of the electromechanical 'bombe' that helped to crack the Enigma cipher; and a sensitive reappraisal of Turing's suspected suicide. A measured portrait at ease with its subject.
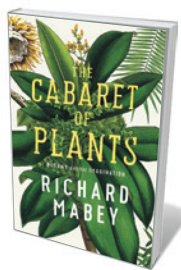
The Cabaret of Plants: Botany and the Imagination
Richard Mabey PROFILE *(2015)*
As a celebrant of the botanical, Richard Mabey has few peers. He is on eloquent form in this portrayal of plants not as dully functional components of natural capital — a "biological proletariat" — but as unruly, autonomous and endlessly fascinating. This engaging scientific and cultural tour takes in ice-age engravings of plant forms; ancients and giants such as bristlecone pines and baobabs; the vast biodiversity of maize (corn); and, as touched on by plant scientist Ian Baldwin (*Nature* **522,** 282–283; 2015), Erasmus Darwin's discovery of "irritability" in *Mimosa pudica* more than 200 years ago.
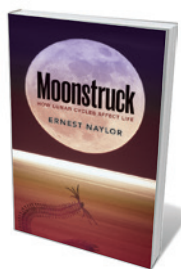
Moonstruck: How Lunar Cycles Affect Life
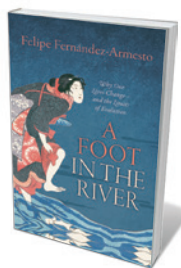*Ernest Naylor* OXFORD UNIVERSITY PRESS *(2015)*
Circadian rhythms are dictated by sunlight and stitched into our genes. But what of the impact of moonlight on life? Marine biologist Ernest Naylor reveals that behavioural patterns linked to lunar phases have been found in animals such as the sea louse *Eurydice*. He also examines Moon-related spawning behaviour in marine species such as grunion and horseshoe crabs, and the sooty tern (*Onychoprion fuscatus*), with its breeding cycle of ten lunar months. For context, Naylor gives us the "full Moon": the deep history, classical science and myth surrounding Earth's beautiful, enigmatic satellite.

The Unknown Universe: What We Don't Know About Time and Space in Ten Chapters
Stuart Clark HEAD OF ZEUS *(2015)*
It is no revelation that some data on the early Universe sit uneasily with the standard model of cosmology. But in his clued-up overview, astronomy journalist Stuart Clark's picture of the yawning gaps in our understanding of the cosmos is fuller than most. Clark tacks back and forth in the history of astronomy, intertwining the discoveries and theories of luminaries from astronomer William Herschel to cosmologist Roger Penrose with speculation on prevailing mysteries such as the nature of dark matter, dark energy and space-time.

A Foot in the River: Why Our Lives Change — and the Limits of Evolution
Felipe Fernández-Armesto OXFORD UNIVERSITY PRESS *(2015)*
Cultural evolutionists paint a partial picture of the speed of change in human cultures, argues historian Felipe Fernández-Armesto. His study, springing from a conference sponsored by the Templeton Foundation, calls for a new interdisciplinarity. He argues that although human culture is born of evolution, it also "changes independently of evolution" because it is a "projection of the human mind" — and its prodigious imaginative capacity. **Barbara Kiser**

Correspondence

US environmentalists must turn out to vote

Nico Stehr rightly argues that democracy is crucial in the fight against global warming, attributing the inadequate response of most democracies to an overall lack of public engagement (*Nature* **525**, 449–450; 2015). Our findings at the Environmental Voter Project indicate that a contributory factor could be a lamentably low turnout by environmentalist voters.

The Environmental Voter Project is a non-partisan, non-profit organization (www.environmentalvoter. org). We estimate that there could be almost 16 million environmentalists in the United States who rarely or never vote — around 7% of the country's voting-eligible population.

To arrive at this figure, we brought in data analysts who used a data-rich voter file to create a national predictive modelling survey to identify people with a very high likelihood of believing that climate change is both human-induced and a crucial issue. They then used public voter files to determine the number in this group who failed to vote in national mid-term elections.

Until many more environmentalists vote, US politicians at least are unlikely to give environmental issues the attention they so badly need.

Nathaniel Stinnett *Environmental Voter Project, Boston, Massachusetts, USA. nathaniel@environmentalvoter.org*

Dutch government appeals climate law

The Dutch government lodged an appeal last month against The Hague District Court's ruling on 24 June that required it to make more-drastic cuts to the country's greenhouse-gas emissions (see K. Purnhagen *Nature* **523,** 410; 2015). The government's appeal seems to be buying time while the courts decide, which demonstrates the weakness of using lawsuits as a policy tool for climate change.

The climate law has been hailed as marking a new era of environmental activism that could spark similar cases in other countries. Critics have warned against undue politicization of the judiciary, which could inhibit nations from entering into binding international commitments.

The Dutch appeal is likely to be based on the government's right to determine policy, on whether the Kyoto Protocol can have such far-reaching effects and on the way the District Court has defined the state's duty of care. Both parties intend to take the case to the Dutch Supreme Court, which could take several years.

Meanwhile, the government is also awaiting the outcome of several studies before launching any policy proposals. This will not be until next summer at the earliest, so the general elections in March 2017 could offer a faster and more effective means of bringing about policy change.

Hanna Schebesta *Wageningen University, the Netherlands; and European University Institute, Florence, Italy.*
Kai Purnhagen *Wageningen University; and Erasmus University Rotterdam, the Netherlands. kai.purnhagen@wur.nl*

Interdisciplinarity: less vague please

The term 'interdisciplinarity' is used to cover a diversity of practices (see *Nature* **525,** 305; 2015). What is crucial for one kind of interdisciplinarity may be immaterial to another.

Without specificity and differentiation, it is impossible to identify factors essential for success. Relevant features include the nature of the problem under investigation; the number of disciplines involved; whether these are closely aligned or disparate; whether the interdisciplinary research is undertaken by an individual or a team; and whether it is engaged with policy and end-user practice (see go.nature.com/ujwu8g).

I investigated one category of interdisciplinary research, in which experts from multiple, diverse disciplines work with end-users on topical problems, to determine the specialist skills required (see go.nature. com/nnsnsx). Synthesizing knowledge, managing remaining unknowns and supporting policy, practice and technological change are all essential. Each of these skills encompasses an array of concepts and methods.

A new 'interdisciplinary' discipline such as 'integration and implementation sciences' can capture, assess and transmit these skills. It could build a college of peer reviewers to improve quality and raise the visibility and influence of interdisciplinarity.

Gabriele Bammer *Australian National University, Acton, Australia. gabriele.bammer@anu.edu.au*

Interdisciplinarity: resources abound

There is growing international consensus on best practice in interdisciplinary research (see *Nature* **525,** 305; 2015). This has been spurred by various online initiatives.

Transdisciplinarity-net, sponsored by the Swiss Academies of Arts and Sciences, offers a toolkit of useful research strategies (see www.transdisciplinarity.ch/ toolbox). The Association for Interdisciplinary Studies provides many resources, including an 'About interdisciplinarity' section that outlines definitions and best practices (see wwwp.oakland. edu/ais). A set of useful short guides is also available (see go.nature.com/faclve) and the Australian I2S site for integration and implementation sciences provides detailed resources (see i2s.anu.edu.au). The Science of Team Science initiative sponsored by the US National Cancer Institute addresses the particular challenges of conducting research in teams (see www.teamsciencetoolkit. cancer.gov).

Notable among the many books on the topic are *Methods for Transdisciplinary Research* (Univ. Chicago Press, 2013) by Matthias Bergmann and colleagues and *Interdisciplinary Research* (Sage, 2011) by Allen Repko.

Because interdisciplinarity is still an emerging approach, such recommendations need reviewing and updating regularly if its potential is to be realized — by those who do interdisciplinary research and by those who study its progress.

Rick Szostak *University of Alberta, Edmonton, Canada. rszostak@ualberta.ca*

Deposited grants buy time in Brazil

Academics who are paralysed by Brazil's political and financial crisis should take heart (see *Nature* **526,** 16–17; 2015). Funds approved for 2014 by the National Council for Scientific and Technological Development, the country's most important funding agency, were fully deposited and are available to principal investigators until 2016 or, in some cases, 2017.

It is crucial, however, that these resources are managed and used wisely. Coordinators must honour their original funding agreements for designated projects.

These guarantees would buy enough time for President Dilma Rousseff to help to restore Brazil's long-standing record of growing and consistent investment in research.

João Ricardo Mendes de Oliveira *Federal University of Pernambuco, Recife, Brazil. joao.ricardo@ufpe.br*

FLUID DYNAMICS

Turbulence spreads like wildfire

A simple model captures the key features of the transition from smooth to turbulent flow for a fluid in a pipe. The findings pave the way for more-complex models and may have engineering ramifications. See Letter p.550

MICHAEL D. GRAHAM

The flow of air over the wing of an aeroplane is smooth and steady at low speeds, as if thin layers of air were sliding over one another. At higher speeds, this laminar flow becomes turbulent — the steady flow gives way to fluctuating eddies that hinder the motion of the aircraft through the air. The processes that drive these fluctuations become self-sustaining during the transition from laminar to turbulent flow, so this transition is a window into the origins of fully developed turbulence. On page 550 of this issue, Barkley et al.[1] present an experimentally validated model of flows in pipes that captures the main features of the transition to the turbulent regime, and illuminates the onset of widespread turbulence.

Every flow is characterized by a dimensionless quantity, known as the Reynolds number, that measures the importance of a fluid's momentum relative to its viscosity. The Reynolds number is low if the fluid's motion is slow, or the spatial extent of the flow is small, or the fluid has large viscosity — any perturbations to the flow will then soon disappear. For a fluid that flows in a pipe, turbulence first occurs when the Reynolds number exceeds about 2,000: a localized perturbation, such as

a jet issuing from the pipe's wall, will evolve into a turbulent puff, a localized patch of turbulence that travels downstream surrounded by regions of laminar flow (Fig. 1). At higher Reynolds numbers, an initially localized perturbation evolves instead into a turbulent pattern that spreads both upstream and downstream relative to the average speed of the fluid in the pipe, leading to what Barkley et al. call the "rise of fully turbulent flow".

Travelling localized patterns arise in many contexts[2] — in forest fires for example. A forest can be described as bistable: that is, it has two steady states, green or burned. A fire is a propagating front that connects these states in space and time by heating up the trees in its path until they catch fire. A large perturbation like a lightning strike is needed to start the fire, but once the fire spreads, the burned forest cannot return to the green state.

A more intricate case is an excitable medium such as a nerve fibre[3]. Here, only one steady state exists: the nerve's relaxed state. Nevertheless, large enough perturbations can drive the nerve fibre to a quasi-steady 'excited' state that can persist for some time before returning to the steady state. In an excitable medium, localized perturbations spread by exciting neighbouring regions, and lead to excitation pulses that are surrounded

by the relaxed state as they travel through the medium.

A mathematical model for patterns in bistable and excitable systems needs two ingredients: one to describe the possible dynamical states at each point in the domain (for example, the presence of two steady states in the bistable case), and the other to describe the communication or transport between points, which is usually modelled as a simple diffusion process. Although they arise in many contexts, models that combine these ingredients are generally called reaction–diffusion models because of the presence of bistability and excitability in chemical reaction processes. These models can reproduce simple propagating fronts or pulses, as well as complex patterns such as spiral waves.

Because the transition of a flow to a turbulent state involves travelling fronts, reaction–diffusion models provide a natural starting point for descriptions of this transition[4]. A turbulent flow is dynamically more complex than a system in a steady or quasi-steady state, but the regime close to the turbulent transition contains simple dynamical states that are akin to turbulent features such as small-scale vortices[5,6]. Diffusion provides some transport of the fluctuations in turbulent flows, but the main transport mechanism is advection, in which the
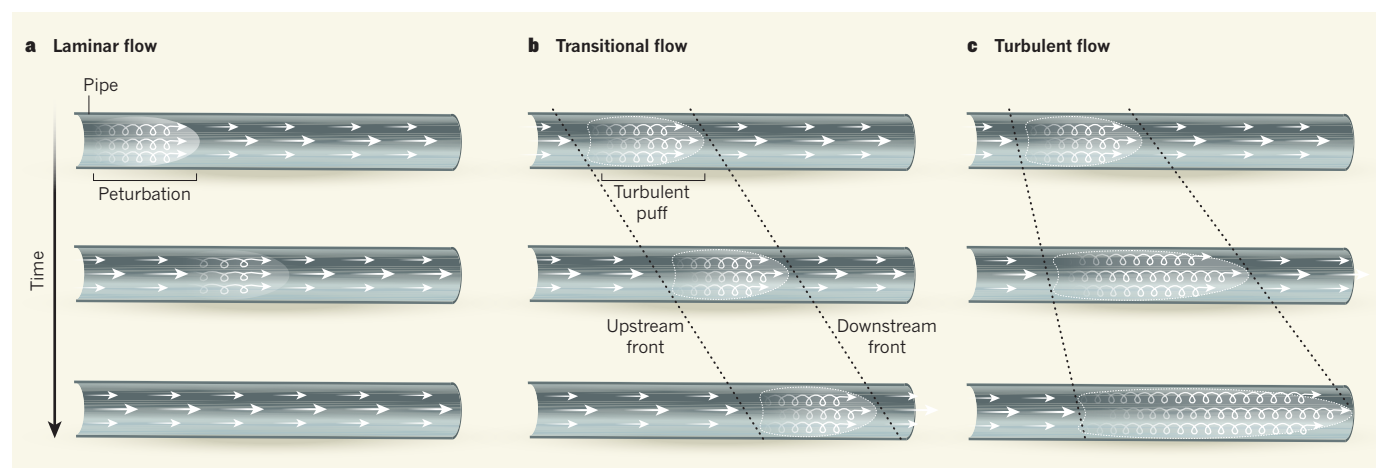

Figure 1 | Pipe flow. Barkley et al.[1] present a mathematical model that captures the evolving characteristics of the flow (arrows) of a fluid in a pipe. **a**, In the laminar regime of smooth and steady flow, externally introduced perturbations soon disappear. **b**, In the transitional regime, perturbations evolve into puffs — turbulent regions whose upstream and downstream fronts propagate at the same speed. The puffs travel without spreading. **c**, In the turbulent regime, a localized perturbation spreads until it overruns the laminar flow that surrounds it.

turbulent flow is carried downstream through bulk motion.

Barkley *et al.* introduce the directionality imparted by advection into a reaction–diffusion model, and use the new model to reproduce in remarkable detail their observations of flows in pipes. In this model there is a laminar steady state that is stable to small perturbations at all Reynolds numbers. However, once the Reynolds number exceeds a threshold value, the system becomes excitable, so that large perturbations of the laminar state evolve into stable localized puffs that are analogous to nerve impulses. As the Reynolds number increases further, the system becomes bistable: initially localized turbulent patches start spreading, first weakly and then strongly, invading the laminar regions until those 'ignite' into turbulence.

Despite having a simple structure, this model quantitatively reproduces the velocities of fronts propagating both upstream and downstream across the Reynolds-number regime that corresponds to the transition from laminar to turbulent flow. Barkley and colleagues thus demonstrate that simple advection-enhanced reaction–diffusion models can capture the large-scale characteristics of a pipe flow's transition regime. This accomplishment highlights the importance of a dynamics-based approach to the understanding of turbulence[6].

The present model has certain limitations that could be addressed in the future. A key simplification is the representation of turbulence by a quasi-steady or steady state. This precludes the model from capturing the internal time and length scales of the fluctuations, which vary between a turbulent region's interior and its edges. Models that can capture more of the dynamics in the turbulent regions could better follow the evolution of turbulence in space and time, and might further elucidate the transition process, including the presence of intermittent laminar flows surrounded by turbulence. Furthermore, the transition and turbulent regimes in polymer solutions[7] or in fluids that contain small particles[8] can be substantially different from the simple fluids described in the current work; extending the experiments and the model to such systems would be of broad interest.

Finally, Barkley and colleagues' study is restricted to flows that have only one extended spatial dimension, in which the travelling patterns are simple pulses. Flows in wide channels or over aircraft wings have an extra extended dimension: in such cases, the transition to the turbulent regime gives rise to complex patterns such as diamond-shaped spots and stripes that are oriented obliquely relative to the main flow direction[9,10]. It would be valuable to develop models for such flows and for various types of perturbations that can trigger turbulent transitions in fluids. This would be of substantial interest to engineers who would like to reduce drag in flows (for example, air drag over wings) by controlling the transition process with suitably designed structures. After all, the best control algorithms are built around a mathematical model of the process to be controlled. ∎

Michael D. Graham *is in the Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.*
e-mail: mdgraham@wisc.edu

1. Barkley, D., Song, B., Mukund, V., Lemoult, G., Avila, M. & Hof, B. *Nature* **526,** 550–553 (2015).
2. Fife, P. C. *Dynamics of Internal Layers and Diffusive Interfaces* (SIAM, 1988).
3. Hodgkin, A. L. & Huxley, A. F. *J. Physiol.* **117,** 500–544 (1952).
4. Pomeau, Y. *Physica D* **23,** 3–11 (1986).
5. Waleffe, F. *J. Fluid Mech.* **435,** 93–102 (2001).
6. Kawahara, G., Uhlmann, M. & van Veen, L. *Annu. Rev. Fluid Mech.* **44,** 203–225 (2012).
7. Graham, M. D. *Phys. Fluids* **26,** 101301 (2014).
8. Matas, J.-P., Morris, J. F. & Guazzelli, É. *Phys. Rev. Lett.* **90,** 014501 (2003).
9. Lemoult, G., Gumowski, K., Aider, J.-L. & Wesfreid, J. E. *Eur. Phys. J. E* **37,** 25 (2014).
10. Duguet, Y. & Schlatter, P. *Phys. Rev. Lett.* **110,** 034502 (2013).

IMMUNOLOGY

Chronic effects of acute infections

Acute infection of mice with an intestinal pathogen leads to long-lasting inflammation that is maintained by intestinal microorganisms. This observation reveals a path by which infection history can affect long-term immune function.

NICOLA HARRIS

Our bodies' history of infections shapes our immune system and can influence the development of subsequent diseases, including inflammatory bowel disease and autoimmune disorders[1]. It has also been postulated[2] that individuals' past infections can undermine vaccine programmes, particularly in developing nations. For certain cases, such as infection with *Streptococcus pyogenes* bacteria and rheumatic heart disease, this link can be explained by the presence of similar antigens (proteins against which the immune system reacts) in both the pathogen and the host[3]. Writing in *Cell*, Morais da Fonseca *et al.*[4] map a different pathway by which infections alter immune status. The authors observe that mice infected with the common intestinal bacterial pathogen *Yersinia pseudotuberculosis* have an altered long-term ability to react to experimental antigens that mimic human exposure to food or oral vaccines.

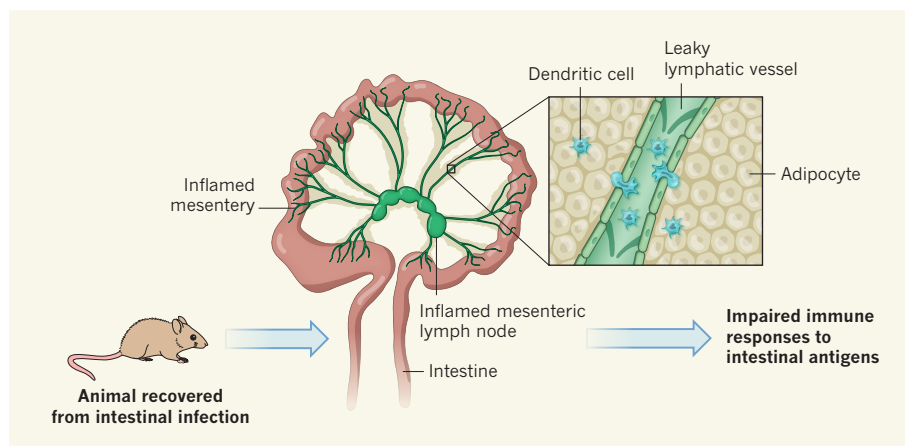Morais da Fonseca and colleagues show

Figure 1 | Leaky lymphatics. Lymphatic vessels in the gut's mesenteric adipose tissue carry dendritic cells (DCs) of the immune system from the intestine to the mesenteric lymph nodes. There, the DCs play a key part in initiating intestinal immune responses, including those that allow the body to tolerate 'foreign' substances (antigens) from food or to respond to oral vaccines. Morais de Fonseca *et al.*[4] show that mice that have recovered from an acute intestinal infection with the bacterium *Yersinia pseudotuberculosis* exhibit persistent inflammation in the mesenteric adipose tissue that results in 'leaky' lymphatics and the loss of migrating DCs into the surrounding tissue. As a consequence, the animals have impaired intestinal immune responses.

that *Y. pseudotuberculosis* infection results in an acute inflammatory episode in the intestine and its associated tissues. However, the surprise came with their observation that, well after the mice had cleared the pathogen, they continued to have swollen lymph nodes (lymphadenopathy) and inflamed mesentery — a fold of tissue that connects the intestine to the body wall. The mesentery is rich in adipocytes (fat cells) and contains lymphatic vessels (which carry fluid and cells from the tissues to the lymph nodes) through which specialized immune cells called dendritic cells (DCs) travel. Although DC migration through mesenteric lymphatic vessels occurs uninterrupted in uninfected mice, the researchers observed that mice that had recovered from *Y. pseudotuberculosis* infection had 'leaky' lymphatic vessels, resulting in the premature exit of DCs and fluids (Fig. 1). This means that these cells failed to complete their journey to the mesenteric lymph nodes, where they are essential for initiating immune responses.

To examine the consequences of persistent lymphatic leakiness in such mice, the authors investigated how these animals responded to previously unencountered intestinal antigens. Using experimental models that mimic oral vaccination or food tolerance (a process by which the immune system prevents reactivity to antigens found in the diet), they found that both types of response were compromised in mice that had previously had an acute infection with *Y. pseudotuberculosis*.

The researchers also observed increased levels of IL-1β and TNFα in the mesenteric adipose tissue following recovery from *Y. pseudotuberculosis* infection. These inflammatory cytokines (proteins involved in intercellular communication) are known to promote lymphatic leakiness[5]. Using germ-free mice, which lack normal resident microbial populations, Morais da Fonseca *et al.* demonstrate that the intestinal microbiota is necessary for the persistent inflammation and lymphatic leakiness after acute infection. This observation prompted the authors to examine whether antibiotic treatment might reverse the negative effects of *Y. pseudotuberculosis* on the host's immune status. Indeed, they found that a short course of antibiotics following recovery from *Y. pseudotuberculosis* infection reduced ongoing mesenteric inflammation and restored immune responsiveness to oral vaccination.

Why and how the intestinal microbiota drives continued mesenteric inflammation after infection has cleared awaits future study. Because mice infected with *Y. pseudotuberculosis* showed no gross long-term differences in the composition of their intestinal microbial communities compared with uninfected animals, it is probable that inflammation resulted from an altered response of the host to the microbiota. In healthy animals, microbial communities are restricted to the intestinal

lumen. However, a variety of injuries or environmental stressors can allow increased numbers of bacteria to cross the intestinal barriers and enter the underlying tissues (a phenomenon called bacterial translocation). Although not experimentally addressed by the researchers, it is possible that *Y. pseudotuberculosis* infection causes long-term changes in intestinal physiology that result in chronic low-level bacterial translocation and entry of these cells into the mesenteric lymphatics. People with inflammatory bowel disease (IBD) display increased bacterial translocation[6], and Morais da Fonseca and colleagues observed leaky lymphatics in a mouse model of IBD. Further investigation is needed to determine whether this process also occurs in humans, and to assess its possible impact on disease severity or immune status.

This work emphasizes the idea that infection history can shape inflammation by disrupting fundamental pathways that are required to initiate immune responses. The unveiling of lymphatic leakiness as one such pathway is an advance in the field that is likely to prompt closer examination of this pathway in diseases that exhibit chronic inflammation, including IBD, autoimmune diseases and obesity. Might we be able to counter the effects of infection history with interventions that improve lymphatic function or restore localized immune responsiveness? Such interventions could aim to resolve chronic inflammation, or to target the intestine to improve barrier

function and prevent bacterial translocation.

Currently, infection history is determined one pathogen at a time, and is based on detecting pathogen-specific antibodies that are made by the immune system in response to pathogen infection and can persist for decades. However, an exciting study this year[7] has reported an antibody-based high-throughput method that uses a single drop of blood to document a person's previous exposure to more than 200 viruses. Expansion of this technology to other pathogens (including bacteria, protozoa and fungi) would allow comparisons between infection history and health status, potentially uncovering links between specific pathogens and diseases. Eventually, the combination of high-throughput screening of infection history and targeted therapy might be used to prevent disease in 'at risk' individuals. ■

Nicola Harris *is in the Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.*
e-mail: nicola.harris@epfl.ch

1. Ercolini, A. M. & Miller, S. D. *Clin. Exp. Immunol.* **155**, 1–15 (2009).
2. Levine, M. M. *BMC Biol.* **8**, 129 (2010).
3. Cunningham, M. W. *et al. J. Immunol.* **141**, 2760–2766 (1988).
4. Morais da Fonseca, D. *et al. Cell* **163**, 354–366 (2015).
5. Cromer, W. E. *et al. Angiogenesis* **17**, 395–406 (2014).
6. Sartor, R. B. *Gastroenterology* **134**, 577–594 (2008).
7. Xu, G. J. *et al. Science* **348**, aaa0698 (2015).

CLIMATE SCIENCE

Small glacier has big effect on sea–level rise

Models of the West Antarctic Ice Sheet predict substantial ice loss over the next few centuries — and that a glacier expected to contribute greatly to sea–level rise may already be unstable.

NATALYA GOMEZ

Sea-level rise is projected to displace communities around the world in the coming centuries. In 2013, the Intergovernmental Panel on Climate Change identified[1] the potential runaway retreat of marine sectors of the West Antarctic Ice Sheet as a major source of uncertainty in predictions of sea-level rise; these sectors contain enough ice to raise average global sea levels by several metres. Writing in *The Cryosphere*, Cornford *et al.*[2] present modelling of the West Antarctic Ice Sheet's response to a warming climate in the next 300 years. Their findings show the potential for substantial retreat, and identify

the Thwaites Glacier — a deep-rooted 'outlet' glacier in the Amundsen Sea Embayment thought to be on the verge of rapid retreat — as probably the largest potential source of sea-level rise over the next few centuries.

Marine sectors of ice in the West Antarctic are kilometres thick and, in some places, sit on bedrock that lies more than a kilometre below the height of the sea surface. These ice sheets gain mass through snowfall from above, flow outwards under the influence of gravity, and lose mass mostly through fast-flowing streams of ice, called outlet glaciers, that feed into ice shelves floating in the surrounding ocean. The grounding line of a marine ice sheet is the zone around the edge of the sheet where the ice is
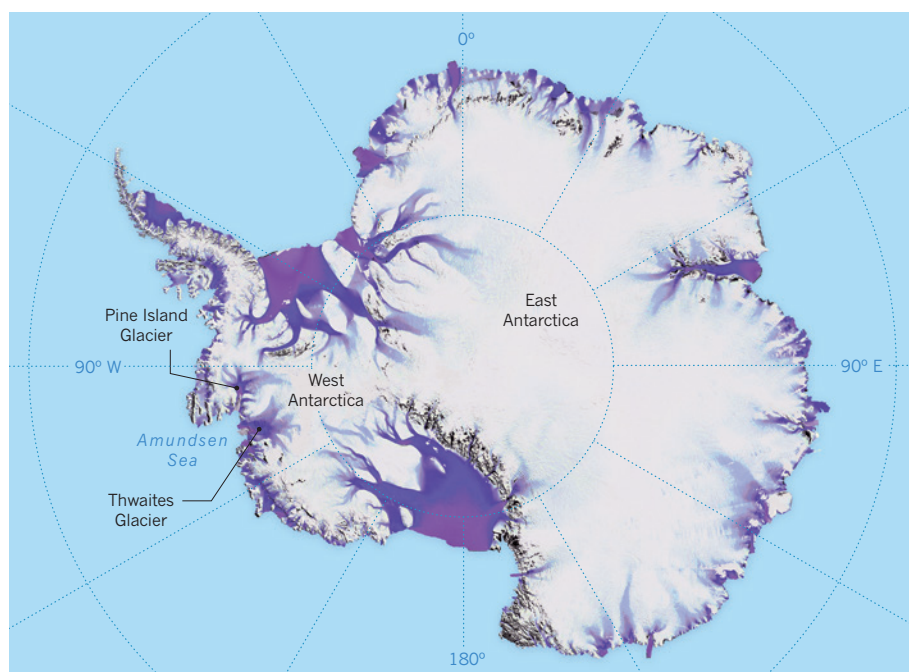
Figure 1 | The Antarctic ice sheet. Marine sectors of ice in Antarctica lose mass mostly through fast-flowing outlet glaciers (purple) that feed into ice shelves in the surrounding ocean. Two of these glaciers — the Thwaites Glacier and the Pine Island Glacier — feed into the Amundsen Sea. Cornford *et al.*[2] have modelled the West Antarctic Ice Sheet's response to climate warming in the next 300 years. They find that the Thwaites Glacier could already be undergoing runaway retreat, and that it may be a substantial source of future, century-timescale sea-level rise. Figure adapted from ref. 12.

just thin enough to float, separating grounded ice sitting on land in the interior from the floating ice shelves.

In a warming climate, marine sectors of ice are thought to be particularly vulnerable. Runaway ice-sheet retreat associated with instability of the grounding line can occur when the bed of a marine outlet glacier deepens upstream of the grounding line[3], as is the case for most of the major outlet glaciers in the West Antarctic[4]. The floating ice shelves on the periphery of these ice sheets stabilize the outlet glaciers, inhibiting the ocean-bound flow of grounded ice and slowing ice loss. But when the ocean warms, these buttressing ice shelves are melted from below and can break up, initiating faster ice flow and rapid retreat of the grounding line[5].

Cornford *et al.* predicted the impact of ongoing climate warming on the stability of the West Antarctic Ice Sheet by using the BISICLES ice-sheet model[6] with a sophisticated treatment of the grounding line, forced by a suite of the most recently available atmosphere and ocean model projections[7,8]. Marine ice sheets interact strongly with both the atmosphere and the ocean, and the computational expense of coupling ice-sheet models to state-of-the-art climate models with a full range of ocean–atmosphere interactions is currently prohibitive. The approximate treatment adopted by the authors is among the best available methods with which to model this complex coupling.

Simulating the migration of the grounding line requires very high spatial resolution, and this limits the spatial and temporal scales that regional ice-sheet models can consider. To clear this technical hurdle, Cornford and colleagues used a numerical approach known as adaptive mesh refinement to focus in on the ice flow at the critical zone near the ice sheet's grounding line. The authors' treatment does not, however, take into account some of the factors that affect ice-sheet dynamics, such as changes in the elevation of Earth's solid surface beneath the ice, and depression of the local sea surface as the gravitational attraction of the ice sheet on the surrounding water weakens[9]. It also neglects some processes that take place at the ice–bed interface, which are challenging to observe. All of the above factors can change the timing and extent of ice-sheet retreat in some regions. But no existing ice-sheet model accounts for all of these effects.

Cornford and colleagues' simulations show that the grounding line retreats by hundreds of kilometres in all major marine outlet glaciers in West Antarctica when ice shelves are melted from below and break apart. However, recent projections using ocean-circulation models[7,8] indicate that ocean warming sufficient to break apart ice shelves in the coming decades will occur only in the Amundsen Sea Embayment, into which the Thwaites and Pine Island glaciers flow (Fig. 1). When the authors used these more-realistic projections of sub-ice-shelf melt rates to drive their simulations, they predicted a contribution of up to 50 mm of global average sea-level rise from the West

Antarctic Ice Sheet by 2100, and up to 150 mm by 2200, the majority of which derives from ice loss in the Amundsen Sea Embayment.

In addition, their results highlight that the stability of the Thwaites Glacier in particular strongly depends on the initial modern state of the ice sheet and bedrock adopted in the simulations. Under a range of reasonable initial conditions, the modelled Thwaites Glacier retreats immediately and rapidly, even without the added forcing of ocean warming breaking up the ice shelf.

These results suggest that strong marine ice-sheet instability may already be under way on Thwaites, even without the help of ice-shelf break-up. This conclusion is supported by other recent investigations into the fate of the Antarctic Ice Sheet, notably a modelling study[10] of the Thwaites Glacier and observation-based work[4] on outlet glaciers in West Antarctica. The emerging picture highlights an urgent need for further observational and modelling explorations of the Amundsen Sea Embayment. Efforts to develop and compare models focused on the future of the Antarctic Ice Sheet in a warming climate are already under way (see ref. 11, for example). ∎

Natalya Gomez *is in the Department of Earth and Planetary Sciences, McGill University, Montreal, Quebec H3A OE8, Canada.*
e-mail: natalya.gomez@mcgill.ca

1. Intergovernmental Panel on Climate Change *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) Summary for Policymakers (Cambridge Univ. Press, 2013).
2. Cornford, S. L. *et al. Cryosphere* **9**, 1579–1600 (2015).
3. Weertman, J. *J. Glaciol.* **13**, 3–11 (1974).
4. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. *Geophys. Res. Lett.* **41**, 3502–3509 (2014).
5. Alley, R. B. *et al. Annu. Rev. Earth Planet. Sci.* **43**, 207–231 (2015).
6. Cornford, S. L. *et al. J. Comput. Phys.* **232**, 529–549 (2013).
7. Timmermann, R. & Hellmer, H. H. *Ocean Dyn.* **63**, 1011–1026 (2013).
8. Hellmer, H. H., Kauker, F., Timmermann, R., Determann, J. & Rae, J. *Nature* **485**, 225–228 (2012).
9. Gomez, N., Pollard, D., & Holland, D. *Nature Commun.* (in the press).
10. Joughin, I., Smith, B. E. & Medley, B. *Science* **344**, 735–738 (2014).
11. Marine Ice Sheet–Ocean Model Intercomparison Project (MISOMIP) and Ice Sheet Model Intercomparison Project 6 (ISMIP6) http://www.climate-cryosphere.org/activities/targeted
12. Rignot, E., Mouginot, J. & Scheuchl, B. *Science* **333**, 1427–1430 (2011).

Kidney tissue grown from induced stem cells

Engineered human cells that can give rise to every cell type have been induced to generate structures that resemble an embryonic kidney. This advance charts a course towards growing transplantable kidneys in culture. SEE LETTER P.564

JAMIE A. DAVIES

Kidney diseases are becoming increasingly common, and there is a shortage of transplantable organs with which to combat them. One solution is to build human kidneys from stem cells. But before this can be done, several difficult problems must be solved. On page 564 of this issue, Takasato *et al.*[1] report taking an important step towards building stem-cell-derived kidneys.

The path from a stem cell to an engineered kidney involves multiple steps. First, stem cells must be persuaded to develop into kidney cells, rather than those of other tissues. Second, once committed to such specialization, the cells must be encouraged to build the intricate, complex anatomy of the kidney. Third, the cultured kidneys must be coaxed to grow and function in a host patient.

Researchers have been making steady progress on the second and third of these steps since 1910, when kidney rudiments were first cultured *in vitro*[2]. Subsequent breakthroughs have enabled both the production of suspensions of fetal animal renogenic (kidney-creating) cells that, in culture, self-organize into small organs and arrangements of tissue called organoids[3–5], and the transplantation of fetal animal kidneys into adult animals[6]. But these advances are of little medical use without conquering the first step: finding a technique for producing renogenic cells, and the vascular progenitors that surround them, from healthy human tissues.

One way to obtain renogenic cells is to grow them from induced pluripotent stem (iPS) cells — adult cells that have been converted in culture to a pluripotent state, from which they can become any cell type in the body[7]. During embryonic development, this process of specialization proceeds through intermediate cell types, and the transition from one stage to the next is triggered by specific signalling proteins such as Wnts or fibroblast growth factors (FGFs). Therefore, biologists who want to direct iPS cells to develop into a particular tissue typically treat them either with signalling molecules or with drugs that mimic these molecules, in an embryologically inspired sequence.

But designing effective protocols for iPS-cell development can be challenging — especially, for example, if the target tissue contains multiple cell types, or if it arises late in embryonic development, many stages away from the pluripotent state. Both of these are true of the kidney, which begins to develop when a human embryo is five weeks old, and which contains cells derived from at least two renogenic progenitor cell types: the ureteric epithelium, which gives rise to collecting ducts that help to maintain the body's balance of fluids and electrolytes, and the metanephrogenic mesenchyme, which matures into nephrons that mediate excretion.

Given the urgent need for kidneys for transplantation, researchers have long been working to obtain renogenic cells from various types of animal and human pluripotent
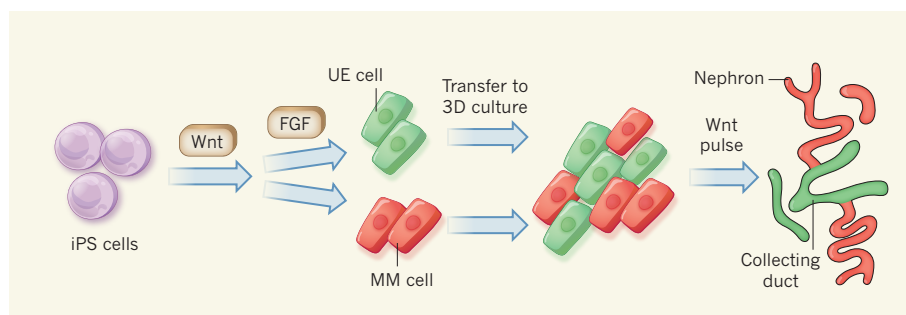

Figure 1 | From stem cell to self-organizing tissue. Takasato *et al.*[1] developed a protocol for growing organized kidney organ buds *in vitro*. Induced pluripotent stem (iPS) cells, which can give rise to any cell type, were exposed to signals from Wnt molecules for four days, and then to molecular FGF signals for five days. Wnt signalling produced a physiological balance between two kidney progenitor cell types: the metanephrogenic mesenchyme (MM) and the ureteric epithelium (UE). The authors transferred these cells to a 3D culture system and then exposed them to another pulse of Wnt signalling, which triggered further development — the cells differentiated and organized themselves into nephrons and pieces of collecting duct that resembled those in an embryonic human kidney.

stem cell. A pioneering study[8] demonstrated that pluripotent cells called mouse embryonic stem cells can be persuaded to express genetic markers of the metanephrogenic mesenchyme and to integrate into host kidneys, albeit with low efficiency. Since then, however, despite incremental improvements, efficiency has continued to be a problem, and too many researchers have relied on marker expression alone as an indicator of success. Yet marker expression is no guarantee that cells will produce a safe, functional tissue — as exemplified by cancer cells that express markers of the tissue from which they arose.

Takasato *et al.* built on previous work to devise a protocol to efficiently turn human iPS cells into metanephrogenic mesenchyme and ureteric epithelium, with the correct balance of cell types (Fig. 1). Their advance was made possible by an improved understanding of the embryonic origin of the two stem-cell types: in particular, the realization that cells that will give rise to ureteric epithelium are exposed only briefly to Wnt signals, whereas those that will give rise to metanephrogenic mesenchyme come from cells that have been exposed for longer. Their protocol thus optimizes the duration of iPS-cell exposure to Wnt-mimicking drugs to produce a balance between the two stem-cell types that approximates the ratio seen *in vivo*. This is followed by exposure to FGF signals, as would occur in a human embryo.

When the researchers cultured the cells as a 3D aggregate, and provided a second, 'trigger' Wnt signal, the metanephrogenic mesenchymal cells developed into nephrons, and the ureteric epithelial cells became collecting ducts. The nephrons matured and produced a sequence of specialized segments that mimicked those in an embryo, along with the connective tissue and vascular progenitor cells that surround embryonic nephrons. Gene expression was comparable with that in first-trimester human fetal-kidney tissue. And the maturing nephrons took up labelled tracer molecules, suggesting that they are functional.

It is vital to emphasize that the result of this process is not a kidney, but an organoid. The structure's fine-scale tissue organization is realistic, but it does not adopt the macro-scale organization of a whole kidney. For example, it is not 'plumbed' into a waste drain, and it lacks large-scale features that are crucial for kidney function, such as a urine-concentrating medulla region containing mature forms of structures called loops of Henle and radially arranged collecting ducts. There is a long way to go until clinically useful transplantable kidneys can be engineered, but Takasato and colleagues' protocol is a valuable step in the right direction.

Even so, these kidney organoids may fulfil a different medical need — the ability to test drug safety on human kidney tissue, rather than in poorly predictive animals[9]. The cell types that are most vulnerable to damage by drugs are present in the organoids, and the authors provide preliminary evidence to demonstrate that the system is indeed damaged by a known renal toxin. It is to be hoped that Takasato *et al.* will team up with toxicologists to perform a full-scale study on the screening potential of their system. The result could be a major step towards animal replacement and improved safety screening for drugs, as well as towards transplantable kidneys. ∎

Jamie A. Davies *is at the Centre for Integrative Physiology, University of Edinburgh, Edinburgh EH8 9XB, UK.*
e-mail: jamie.davies@ed.ac.uk

1. Takasato, M. *et al. Nature* **526,** 564–568 (2015).
2. Carrel, A. & Burrows, M. *J. Am. Med. Assoc.* **55,** 2057–2058 (1910).
3. Unbekandt, M. & Davies, J. A. *Kidney Int.* **77,** 407–416 (2010).
4. Lusis, M., Li, J., Ineson, J., Christensen, M. E., Rice, A. & Little, M. H. *Stem Cell Res.* **5,** 23–39 (2010).
5. Lawrence, M. L., Chang, C. H. & Davies, J. A. *Sci. Rep.* **13,** 9092 (2015).
6. Rogers, S. A., Lowell, J. A., Hammerman, N. A. & Hammerman, M. R. *Kidney Int.* **54,** 27–37 (1998).
7. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
8. Kim, D. & Dressler, G. R. *J. Am. Soc. Nephrol.* **16,** 3527–3534 (2005).
9. Fletcher, A. P. *J. R. Soc. Med.* **71,** 693–696 (1978).

MICROBIOLOGY

Conductive consortia

Physiological analyses, electron microscopy and single-cell chemical imaging suggest that direct electron transfer occurs between the members of methane-oxidizing microbial consortia. SEE ARTICLE P.531 AND LETTER P.587

MICHAEL WAGNER

Emission of the greenhouse gas methane from the seabed is controlled by its anaerobic oxidation coupled with sulfate reduction. This globally important process, which consumes most of the methane released and thus regulates the climate on our planet, is often mediated by dense aggregates of two specialized microorganisms: anaerobic methanotrophic archaea and deltaproteobacteria. These bi-species consortia were discovered about 15 years ago[1], but how the partners make a living from this low-energy-yielding process has been a mystery, although it has repeatedly been speculated that exchange of a diffusible metabolite between the partner microbes is essential[2–4]. Two papers in this issue challenge this hypothesis. McGlynn *et al.*[5] (page 531) demonstrate that the relationship between species intermixing and cellular activity patterns is inconsistent with transfer of a diffusible compound, and Wegener *et al.*[6] (page 587) show with physiological experiments that transfer of intermediates cannot explain the growth of the aggregates in culture.

Within the consortia, the anaerobic methanotrophic archaea (ANME) oxidize methane to carbon dioxide by reversing the classical pathway of methanogenesis[7]. This process is energetically favourable only if the resulting electrons are efficiently transferred to sulfate. Several hypotheses have been put forward to explain how this is achieved (Fig. 1). One theory proposes[2–4] that electrons are transferred from the archaea to their deltaproteobacterial sulfate-reducing partners through the production and consumption of a diffusible metabolite, such as hydrogen, formate or methanethiol. This model is consistent with long-known strategies for electron transfer between microbial species, but experimental evidence that it occurs in this system is lacking.

An alternative model[8], based on a wide array of experimental data, predicts that ANME can autonomously perform anaerobic methane oxidation by reducing sulfate (SO_4^{2-}) to zero-valent sulfur (S^0). This reacts with environmental sulfide to form sulfur compounds that are used by the deltaproteobacteria, which act as commensal organisms — they benefit from the relationship without affecting the archaea. Although this model can explain the frequently observed occurrence of ANME in the environment without deltaproteobacteria[9], the enzymatic machinery for sulfate reduction in ANME has not yet been discovered.

McGlynn *et al.* now use chemical imaging and isotope labelling to measure the metabolic activity of single cells in 62 consortia from deep-sea sediments at an active methane source, and show that cellular activity patterns are independent of the distance between partner cells. Furthermore, they find that the activities of entire aggregates are not related to the spatial distribution of the microbial members.

These findings are in marked contrast to theoretical predictions on activity patterns in multispecies consortia that are driven by the exchange of microbial metabolites. According to such predictions, interacting species in close contact with each other would be more active than those separated by greater distances, and thus well-intermixed aggregates should be

more active than poorly intermixed consortia. And in the alternative model, in which ANME reduce the sulfate and feed sulfur compounds to the deltaproteobacteria[8], one would expect archaeal activity to be unrelated to bacterial activity, but that bacterial activity would still be higher the closer the bacterial cells are to the archaeal cells. This pattern was also not observed by McGlynn and colleagues.

Wegener *et al.* tested how the *in vitro* activity of consortia microbes enriched from a deep-sea sediment responded to the addition of various organic and inorganic compounds (including zero-valent sulfur) that could be produced by the ANME and consumed by the bacteria. Only the addition of hydrogen stimulated sulfate reduction in the cultures, suggesting that this is the only compound suitable for shuttling electrons to the deltaproteobacteria. Successful cultivation of the deltaproteobacteria alone with hydrogen as the only energy source confirmed this finding. However, the authors report that the level of hydrogen produced by the consortia when sulfate reduction is inhibited is too low to explain the growth of the deltaproteobacteria. Thus, although hydrogen may be used as a growth substrate for the deltaproteobacteria, it is clearly not the driver of anaerobic methane oxidation by the consortia.

But if interspecies metabolite exchange is not happening in these consortia, how might they jointly generate energy? Stimulated by previous speculations[3], McGlynn *et al.* and Wegener *et al.* hypothesized that direct interspecies electron transfer (DIET)[10] occurs between ANME and their bacterial partner through electrical connections. This trick would enable the deltaproteobacteria to efficiently reduce sulfate through using the electrons generated by the archaea from methane oxidation. McGlynn and colleagues present modelling data predicting that, in aggregates powered by DIET, there would be little correlation between the activity and the spatial distribution of cells of the two partner species, more in line with their experimental data.

DIET has recently been recognized as an alternative to interspecies transfer of diffusible intermediates such as hydrogen and formate[10], and methane-producing microorganisms have been shown to accept electrons by this mechanism[11]. By analysing available genomic information for ANME and the deltaproteobacteria, the two research groups discovered genes encoding large multi-haem cytochromes (proteins that mediate electron transport) and type IV pili (cellular appendages), which are
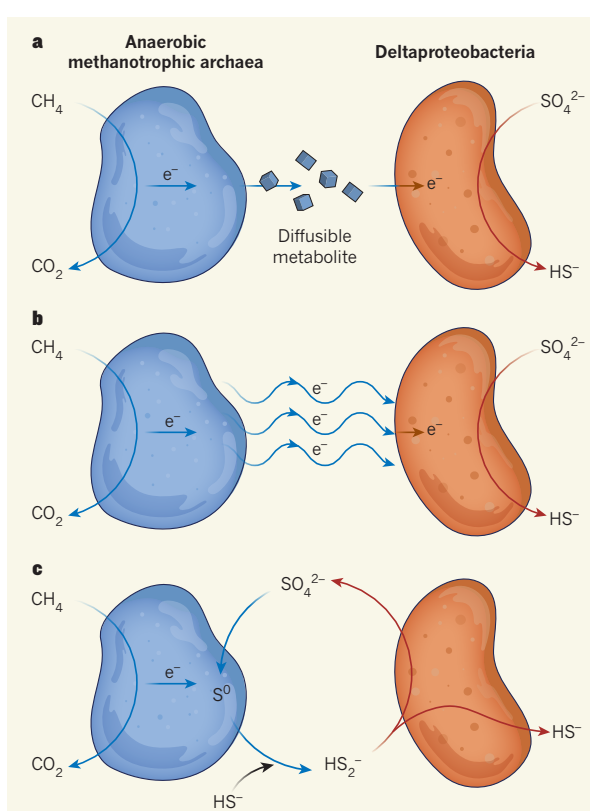
Figure 1 | Interaction modes. Anaerobic methanotrophic archaea (ANME) and deltaproteobacteria are marine microorganisms that together mediate the anaerobic oxidation of methane coupled with the reduction of sulfate. Three strategies for this interaction have been postulated. **a**, One proposes[2–4] that ANME oxidize methane (CH_4) to carbon dioxide and transfer the electrons (e^-) obtained to deltaproteobacteria through diffusible metabolites such as hydrogen, formate or methanethiol. The bacteria use the electrons to reduce sulfate (SO_4^{2-}) to sulfide (HS^-). **b**, McGlynn *et al.*[5] and Wegener *et al.*[6] suggest that the same oxidation and reduction reactions take place, but that direct electron transfer occurs by means of electrical connections involving multi-haem cytochrome proteins and appendages called type IV pili. **c**, An alternative hypothesis[8] is that ANME oxidize methane but also reduce sulfate to zero-valent sulfur (S^0), which, after release, forms disulfide (HS_2^-) in the presence of environmental sulfide. The deltaproteobacteria convert the disulfide to sulfide and sulfate, but, in contrast to the interactions in **a** and **b**, this activity is not necessary for powering anaerobic methane oxidation.

both hallmarks of organisms capable of extracellular electron transfer. Consistent with this finding, McGlynn and colleagues show that the extracellular space in the aggregates could be stained with a cytochrome-reactive compound, suggesting that conductive haem proteins are present between the ANME and the bacterial cells. Wegener *et al.* even managed to directly visualize nanowire-like structures, 10 nanometres thick and up to 1 mm long, that connect the ANME and deltaproteobacterial cells. These structures were not seen on the surface of the deltaproteobacteria in cultures without ANME, lending further support to their role in connecting the two partners.

The results presented in these two papers are a major advance in microbiologists' struggle to decipher the enigmatic metabolism of anaerobic-methane-oxidizing consortia, and

they provide the first experimental support for DIET in these assemblages. One of the biggest remaining challenges is to prove this mode of interaction unequivocally. Can ANME grow on anodes (electron-accepting electrodes) without their bacterial partners, and if so, what mechanism do they use for electron transfer? Members of the consortia cannot yet be cultured in isolation or genetically manipulated, so it will not be a straightforward matter to further characterize the role of the multi-haem cytochromes and the type IV pili. A logical next experiment would be to identify the location of these molecules in the aggregates. Furthermore, it will be fascinating to investigate whether DIET enables partnerships of ANME with other microbes and provides access to electron acceptors other than sulfate.

With these findings, anaerobic methane oxidation becomes another hot candidate for the increasing number of processes recognized as being driven by electromicrobiology[12]. A pressing task now is to determine how widespread DIET is in the various known ANME and deltaproteobacterial partner lineages. Interaction strategies between microorganisms can be subject to rapid evolution, and it is thus conceivable that phylogenetically identical or closely related partner species use different modes of interaction[13]. ∎

Michael Wagner *is in the Department of Microbiology and Ecosystem Science and the Research Network 'Chemistry meets Microbiology', University of Vienna, Vienna 1090, Austria.*
e-mail: wagner@microbial-ecology.net

1. Boetius, A. *et al. Nature* **407,** 623–626 (2000).
2. Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. *Global Biogeochem. Cycles* **8,** 451–463 (1994).
3. Meyerdierks, A. *et al. Environ. Microbiol.* **12,** 422–439 (2010).
4. Moran, J. J. *et al. Environ. Microbiol.* **10,** 162–173 (2008).
5. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. *Nature* **526,** 531–535 (2015).
6. Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. *Nature* **526,** 587–590 (2015).
7. Hallam, S. J. *et al. Science* **305,** 1457–1462 (2004).
8. Milucka, J. *et al. Nature* **491,** 541–546 (2012).
9. Knittel, K. & Boetius, A. *Annu. Rev. Microbiol.* **63,** 311–334 (2009).
10. Shrestha, P. M. & Rotaru, A. E. *Front. Microbiol.* **5,** 237 (2014).
11. Rotaru, A.-E. *et al. Appl. Environ. Microbiol.* **80,** 4599–4605 (2014).
12. Lovley, D. R. *Annu. Rev. Microbiol.* **66,** 391–409 (2012).
13. Summers, Z. M. *et al. Science* **330,** 1413–1415 (2010).

ECOLOGY

Mangrove maintenance

The stilt-rooted trees of mangrove forests host rich biological diversity, as well as supporting fisheries and protecting shores from storm damage and erosion. These tidal-zone trees can maintain an appropriate soil elevation for local sea levels and inundation rates by accreting sediment or organic material around their roots (pictured, mangroves in Indonesia). But on page 559 of this issue, Lovelock *et al.* (C. E. Lovelock *et al. Nature* **526,** 559–563; 2015) show that for many forests, current rates of sea-level rise outpace this adaptive capacity.

Assessing 27 sites across the Indo-Pacific, the authors find that sediment availability is a key survival factor for mangroves in the region. But river damming and land-use change are reducing sediment supply. The researchers' modelling predicts that, at current rates of sea-level rise, many mangrove forests could be submerged by 2070. **Marian Turner**

EXOPLANETS

A glimpse of Earth's fate

Analysis of data from the Kepler space observatory and ground–based telescopes has led to the detection of one, and possibly several, minor planets that are in a state of disintegration in orbit around a white dwarf star. SEE LETTER P.546

FRANCESCA FAEDI

What will Earth's fate be when the Sun dies? Writing on page 546 of this issue, Vanderburg *et al.*[1] offer a dramatic answer in their discovery of one, or possibly several, dying minor planets, which have a rocky bulk and chemical composition similar to Earth's. These small bodies are in a tight orbit around a white dwarf, the remnant of a Sun-like star that has reached the end of its life, and they are being shredded to pieces by the star's strong gravity and radiation field. The authors' observations of the system reveal multiple transit signals — the periodic dimming of the stellar light caused by passing foreground objects — induced by one or several disintegrating bodies that have orbital periods of 4.5 to 4.9 hours.

The vast majority of exoplanets discovered up to now orbit main-sequence stars, which,

like our Sun, are in the prime of their lives. By contrast, the minor planets discovered by Vanderburg *et al.* are orbiting a member of the stellar graveyard, but not for much longer. A star like the Sun reaches the end of its life when the nuclear fuel in the stellar core is exhausted. During this process the Sun will expand and become a red-giant star that will engulf the inner planets Mercury and Venus. Whether the Earth will be swallowed up by the bloated Sun is still a matter of debate; however, even if the Earth survives, its surface will be roasted. Following the red-giant phase, and before becoming a white dwarf, the Sun will lose a large fraction of its original mass.

This overall process will destabilize the planetary orbits[2] and might cause collisions between the planets, similar to those that occurred during the infancy of the Solar System. Some planets might in this way be shattered to pieces resembling asteroids. If

such an asteroid wanders too close to the white dwarf it will be ripped apart by strong tidal (gravitational) forces, and a circumstellar dust disk will form[3,4] of similar chemical composition to that of the original planetary core. Such a disk can be then accreted onto the atmosphere of the white dwarf[5].

White dwarfs are small but extremely dense, and so they have strong gravitational fields. Consequently, elements heavier than helium (called metals by astronomers) that fall into a white dwarf's pure hydrogen or helium atmosphere are expected to sink towards the star's core within a matter of days. But astronomers have discovered metals such as carbon, silicon, oxygen and iron in the atmospheres of one-third of all known white dwarfs[6], and observations at infrared wavelengths have revealed that some of these stars have circumstellar dust disks[7,8]. Thus the atmospheric pollution of white dwarfs by metals such as these, which plausibly originated in circumstellar disks, provides strong evidence that a substantial fraction of white dwarfs have devoured broken-up planets or asteroids of chemical compositions similar to those of terrestrial bodies. After all, carbon, silicon, oxygen and iron make up roughly 93% of Earth's mass[9].

The NASA Kepler space observatory was launched in 2009 and since then it has been obtaining high-precision photometric measurements of the brightness of stars in the constellations Cygnus and Lyra[10]. By monitoring planetary transits, these observations
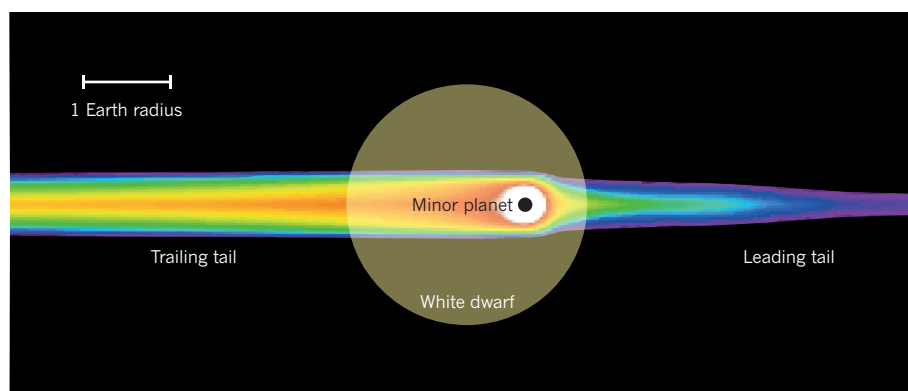
Figure 1 | A minor planet transits a white dwarf. Vanderburg *et al.*[1] analysed data from the white dwarf star WD 1145+017 taken by the Kepler space observatory and ground-based telescopes. The data revealed transit features (periodic attenuations of the stellar brightness) that are best explained by the passage in front of the star of one or several disintegrating minor planets. The authors performed simulations of this process in which a minor planet that orbits very close to the star loses mass in the form of dust particles that generate leading and trailing cometary tails (colours indicate dust density). The various phases of the transit (leading tail, minor planet core and trailing tail) induce attenuations of the stellar brightness of different magnitudes and durations. Adapted from Fig. S7 of ref. 1.

have led to the detection of hundreds of rocky exoplanets, and revolutionized this research field. Following technical problems during the first part of the mission, a second mission[11] (dubbed K2) was planned that included targets such as white dwarfs.

Vanderburg *et al.* analysed photometric observations of WD 1145+017, a white dwarf that was observed during K2, and discovered multiple transit features in the data. The authors then used an established statistical regression method[12], and identified transit signals induced by bodies that have orbital periods of 4.5 to 4.9 hours. These transits immediately seemed peculiar: they were shallow, which means that they did not cause strong dimming of the star's brightness, and they lasted an unusually long time — about 40 to 80 minutes. Because white dwarfs are small (about the size of Earth), a solid body that passes in front of the stellar disk is expected to induce a short transit event lasting only a minute or so[13,14].

To clarify the nature of the transits, Vanderburg and colleagues further observed the system using several ground-based telescopes. This additional photometric monitoring revealed very deep (40% of the stellar light was blocked), short-duration (5-minute) asymmetric transits separated by the dominant 4.5-hour period identified in the K2 data. The typical geometry of a transit dictates that a small spherical disk passing in front of a larger spherical disk would yield a perfectly symmetrical transit light curve. These observations, however, obfuscated interpretations even more, because the transit signals seemed to be not only different, but also out of phase with those seen in the Kepler data, and morphologically asymmetrical between the initial and final parts of the transit. Finally, Vanderburg *et al.* analysed spectra of the white dwarf and detected metals in its atmosphere.

As the authors state, a possible explanation of the unusual transit events is one or several minor planets in orbit around WD 1145+017 that are losing material into space as they break into pieces. The evaporated material is expelled in a wind, forming a cloud of molecules that condenses behind the disintegrating body in the shape of a cometary tail (Fig. 1). Evaporating planets have been observed transiting main-sequence stars[15,16], and those observations showed asymmetrical transit profiles and variable transit depths like the ones in the current study. In all these cases, a dust cloud trailing the evaporating body can explain the transits' variable depths, asymmetrical profiles and unusually long durations.

It is extremely exciting that astronomers

have recorded the final throes of a planetary system, and further analysis of its properties is warranted. This research will have a transformative impact on the exoplanet field and will stimulate studies of the chemistry of planetary interiors. Future observations of evaporating planets and metal-polluted white dwarfs might even allow scientists to distinguish between material that originated in a planet's core as opposed to its mantle. Although Earth's final days are a long way into the future, this research has allowed us a glimpse of the probably inescapable outcome. ∎

Francesca Faedi *is in the Department of Physics, University of Warwick, Coventry CV4 7AL, UK.*
e-mail: f.faedi@warwick.ac.uk

1. Vanderburg, A. *et al.* Nature **526,** 546–549 (2015).
2. Debes, J. H. & Sigurdsson, S. *Astrophys. J.* **572,** 556 (2002).
3. Jura, M. A. *Astrophys. J. Lett.* **584,** L91 (2003).
4. Veras, D., Leinhardt, Z. M., Bonsor, A. & Gänsicke, B. T. *Mon. Not. R. Astron. Soc.* **445,** 2244–2255 (2014).
5. Jura, M. *Astrophys. J.* **584,** L91 (2003).
6. Zuckerman, B., Melis, C., Klein, B., Koester, D. & Jura, M. *Astrophys. J.* **722,** 725 (2010).
7. Farihi, J., Jura, M. & Zuckerman, B. *Astrophys. J.* **694,** 805 (2009).
8. Barber, S. D. *et al. Astrophys. J.* **760,** 26 (2012).
9. Farihi, J., Gänsicke, B. T. & Koester, D. *Science* **342,** 218–220 (2013).
10. Borucki, W. J. *et al. Science* **327,** 977–980 (2010).
11. Howell, S. B. *et al. Publ Astron. Soc. Pac.* **126,** 398 (2014).
12. Kovács, G., Zucker, S. & Mazeh, T. *Astron. Astrophys.* **391,** 369–377 (2002).
13. Faedi F., West, R. G., Burleigh, M. R., Goad, M. R. & Hebb, L. *Mon. Not. R. Astron. Soc.* **410,** 899–911 (2011).
14. Agol, E. *Astrophys. J.* **731,** L31 (2011).
15. Rappaport, S. *et al. Astrophys. J.* **752,** 1 (2012).
16. Sanchis-Ojeda, R. *et al.* Preprint at http://arxiv.org/abs/1504.04379 (2015).

EVOLUTION

An avian explosion

The genome sequences of 198 bird species provide an unprecedented combination of breadth and depth of data, and allow the most robust resolution so far of the early evolutionary relationships of modern birds. SEE LETTER P.569

GAVIN H. THOMAS

The fossil record offers a compelling narrative of avian evolution. There are few known fossils of modern birds from the Cretaceous period (around 145 million to 66 million years ago), but most major modern-bird lineages are well represented in fossils from the Palaeogene (around 66 million to 23 million years ago). It is suggested that, following the mass extinction at the end of the Cretaceous that famously wiped out the non-avian dinosaurs, birds went through

a (geologically) brief recovery followed by an explosive species radiation. But agreement between the fossil record and phylogenetic trees has been conspicuously absent. On page 569 of this issue, Prum *et al.*[1] use genome sequences of 198 bird species spanning the entire radiation of modern birds to build a new phylogeny. The authors' tree resolves the branching order at the origins of modern birds and strongly supports a rapid radiation of major bird lineages soon after the Cretaceous–Palaeogene mass extinction.

The evolutionary relationships between

modern birds have been notoriously difficult to resolve[2–4]. The problem for evolutionary ornithologists is that ancient divergences over short periods are exceedingly difficult to tease apart, and this has limited their ability to make robust inferences about early bird evolution. Prum and colleagues use a genomic sequencing technique called anchored hybrid enrichment[5] to sample highly conserved (slowly evolving) regions of the genome and faster-evolving flanking regions that together are particularly well suited to teasing apart rapid, but ancient, radiations.

Prum and colleagues' phylogeny differs dramatically from another analysis reported last year, by Jarvis et al.[2], of an exceptionally large data set of more than 40 million base pairs of nucleotide sequence data from 48 avian genomes. Not surprisingly, the conflict is focused on the earliest branching events that separate major non-passerine taxa (Fig. 1). For example, a clade including hummingbirds, swifts and nightjars is shown to be sister to the rest of the Neoaves — a clade that includes all living bird species except for Palaeognathae (such as ostriches and kiwis), Galliformes (landfowl) and Anseriformes (ducks and geese) — rather than sister to grebes and flamingos. And an entirely new clade, called Aequorlitornithes, is identified with strong support and consists of the majority of Neo-avian groups of waterbirds.

Why does the topology of some parts of the two trees differ so fundamentally despite the use of exceptionally large genomic data sets in both studies? One possibility is that an explosion of speciation after the Cretaceous–Palaeogene extinction saw all major lineages of birds branch off near-simultaneously. Indeed, the early diversification of birds may have been so rapid that it resembles a network, or bush, rather than a beautifully bifurcating tree of life. Recent support for this idea comes from the finding[6] that a process called incomplete lineage sorting (ILS) was rampant when the major lineages of birds diversified. The effect of ILS is that different parts of the genome yield different evolutionary relationships and produce a pattern akin to tangled roots, rather than a tree. ILS is usually identified only in recent species radiations, but large genomic data sets allow for detailed tests that delve deeper into the evolutionary past. Although the difference between the phylogenies could simply result from the two data sets having sampled different parts of the genome that happen to be incongruent as a result of ILS, this seems unlikely given the vast amount of genomic data sequenced in each study.

Instead, Prum et al. explore an alternative and perhaps more likely cause for discrepancies between phylogenetic hypotheses — a phenomenon called long-branch attraction. Long-branch attraction occurs when distant evolutionary relatives are incorrectly inferred to be close relatives; this can arise if evolution
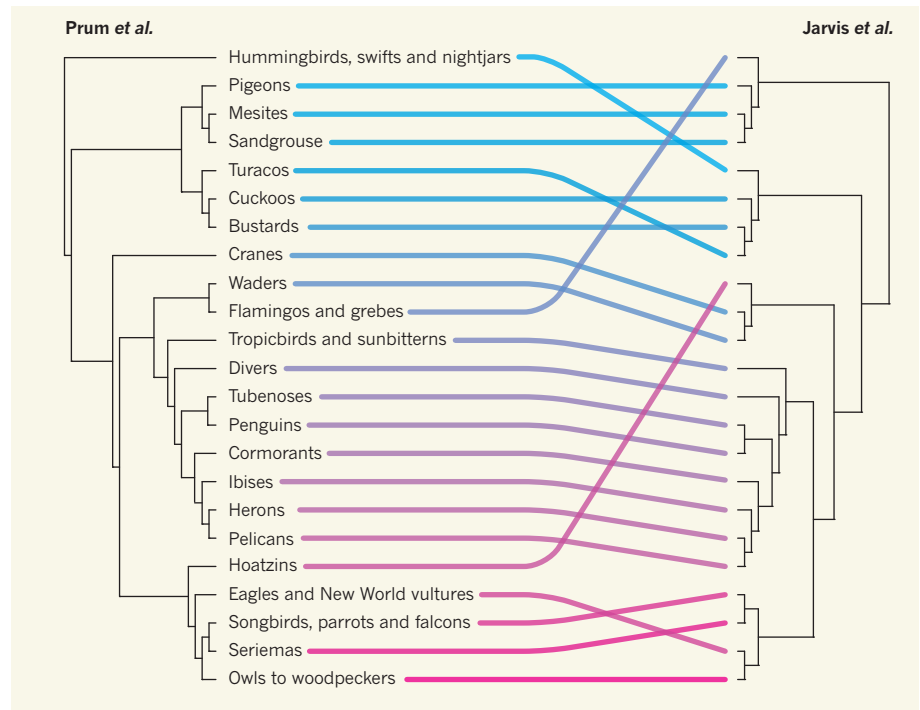
Figure 1 | A comparison of avian phylogenies. The phylogenetic relationships presented by Prum et al.[1] and Jarvis et al.[2] have here been distilled down to major bird lineages. The comparison reveals several differences between the postulated evolutionary relationships of taxa.

has proceeded at an exceptionally high rate, or when one lineage has no close relatives. Prum et al. deal with long-branch attraction by sampling both deeply (the number of nucleotides sequenced) and broadly (the number of species). This strategy is rooted in well-established systematic theory, which shows that sampling more species can break long branches. When the 198-species data set of Prum et al. is reduced to include only the 48 species in Jarvis and colleagues' data, Prum and colleagues' phylogeny breaks down because many relationships change fundamentally. The breadth of sampling is critical.

A well-resolved phylogeny is the basis of robust dating. Congruence between dates from molecular phylogenies and the fossil record is a rare thing, and for birds it is likely to prove controversial[7,8]. Prum and colleagues' conclusion of an explosive radiation after the Cretaceous–Palaeogene mass extinction is markedly different from the conclusions of many previous molecular studies, which typically suggest that most major avian orders and many families originated further back in the Cretaceous period (for examples, see refs 9–11).

Although genomic-scale data can add precision to dating estimates, accuracy relies on the quality of the fossil material and our ability to place it in the correct evolutionary context. Phylogenetic trees are calibrated using the fossil record with all its inherent imperfections. But the fossil record of birds is patchy, incomplete and geographically biased. For most divergences no fossil evidence is available, and for others the fossil record probably underestimates the true age of origination, because the fossils discovered for a particular group are likely to be younger than the age of divergence. The age of the root of the avian tree is particularly contentious, with different calibrations placing the explosive radiation of Neoaves either before[7] or after[8] the Cretaceous–Palaeogene mass extinction. In the absence of a perfect fossil record, the best we can do is experiment with different calibration dates and levels of uncertainty around those dates. The new genomic data sets await extensive experimentation of this type, and so, as compelling as the historical narrative may be, it is perhaps for now best treated with hopeful caution. ∎

Gavin H. Thomas is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.
e-mail: gavin.thomas@sheffield.ac.uk

1. Prum, R. O. et al. Nature **526**, 569–573 (2015).
2. Jarvis, E. D. et al. Science **346**, 1320–1331 (2014).
3. Hackett, S. J. et al. Science **320**, 1763–1768 (2008).
4. McCormack, J. E. et al. PLoS ONE **8**, e54848 (2013).
5. Lemmon, A. R., Emme, S. & Lemmon, E. C. System. Biol. **61**, 721–744 (2012).
6. Suh, A., Smeds, L. & Ellegren, H. PLoS Biol. **13**, e1002224 (2015).
7. Mitchell, K. J., Cooper, A. & Phillips, M. J. Science **349**, 1460 (2015).
8. Cracraft, J. et al. Science **349**, 1460 (2015).
9. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. Nature **491**, 444–448 (2012).
10. Pacheco, M. A. et al. Mol. Biol. Evol. **28**, 1927–1942 (2011).
11. Brown, J. W., Rest, J. S., García-Moreno, J., Sorenson, M. D. & Mindell, D. P. BMC Biol. **6**, 6 (2008).

This article was published online on 7 October 2015.

ARTICLE

Non-coding recurrent mutations in chronic lymphocytic leukaemia

Xose S. Puente[1], Silvia Beà[2], Rafael Valdés-Mas[1], Neus Villamor[3], Jesús Gutiérrez-Abril[1], José I. Martín-Subero[4], Marta Munar[5], Carlota Rubio-Pérez[6], Pedro Jares[7], Marta Aymerich[3], Tycho Baumann[8], Renée Beekman[2], Laura Belver[9], Anna Carrio[3], Giancarlo Castellano[2], Guillem Clot[2], Enrique Colado[10], Dolors Colomer[3], Dolors Costa[3], Julio Delgado[8], Anna Enjuanes[7], Xavier Estivill[11], Adolfo A. Ferrando[9], Josep L. Gelpí[5], Blanca González[3], Santiago González[5], Marcos González[12], Marta Gut[13], Jesús M. Hernández-Rivas[12], Mónica López-Guerra[3], David Martín-García[2], Alba Navarro[2], Pilar Nicolás[14], Modesto Orozco[5], Ángel R. Payer[10], Magda Pinyol[7], David G. Pisano[15], Diana A. Puente[1], Ana C. Queirós[4], Víctor Quesada[1], Carlos M. Romeo-Casabona[14], Cristina Royo[2], Romina Royo[5], María Rozman[3], Nuria Russiñol[2], Itziar Salaverría[2], Kostas Stamatopoulos[16], Hendrik G. Stunnenberg[17], David Tamborero[6], María J. Terol[18], Alfonso Valencia[15], Nuria López-Bigas[6], David Torrents[5], Ivo Gut[13], Armando López-Guillermo[8], Carlos López-Otín[1]§ & Elías Campo[3]§

Chronic lymphocytic leukaemia (CLL) is a frequent disease in which the genetic alterations determining the clinicobiological behaviour are not fully understood. Here we describe a comprehensive evaluation of the genomic landscape of 452 CLL cases and 54 patients with monoclonal B-lymphocytosis, a precursor disorder. We extend the number of CLL driver alterations, including changes in *ZNF292*, *ZMYM3*, *ARID1A* and *PTPN11*. We also identify novel recurrent mutations in non-coding regions, including the 3′ region of *NOTCH1*, which cause aberrant splicing events, increase NOTCH1 activity and result in a more aggressive disease. In addition, mutations in an enhancer located on chromosome 9p13 result in reduced expression of the B-cell-specific transcription factor PAX5. The accumulative number of driver alterations (0 to ≥4) discriminated between patients with differences in clinical behaviour. This study provides an integrated portrait of the CLL genomic landscape, identifies new recurrent driver mutations of the disease, and suggests clinical interventions that may improve the management of this neoplasia.

CLL is a B-cell neoplasia that exhibits a very heterogeneous course, with some patients following an indolent disease course, clearly contrasting with others experiencing an aggressive disease[1–3]. Patients have been classically categorized in two groups, depending on whether their tumour B cells express B-cell receptor (BCR) immunoglobulin with immunoglobulin heavy variable (IGHV) genes bearing somatic hypermutation (IGHV-mutated) or not (IGHV-unmutated)[4]. Further studies have led to the identification of additional biological features with prognostic value for CLL patients[5–8]. However, the molecular mechanisms responsible for the initiation and heterogeneous evolution of CLL remain largely unknown.

Whole-genome sequencing (WGS) and whole-exome sequencing (WES) studies in CLL patients have identified recurrently mutated genes such as *NOTCH1*, *SF3B1*, *TP53*, *BIRC3* and *POT1*, and delineated clonal evolution events in this neoplasia[9–15]. Moreover, recent works have profiled the transcriptome and the DNA methylome of many CLL cases[16–18]. Nevertheless, these studies have unveiled a high level of molecular heterogeneity, thus creating the need for integrated analysis of different genomic parameters in a larger number of patients. In this work, and as part of the International Cancer Genome Consortium (ICGC) project[19], we have performed a comprehensive analysis of the genetic alterations driving the oncogenic transformation in 506 patients with monoclonal B-lymphocytosis (MBL) or CLL. We have also carried out additional genomic studies involving single nucleotide polymorphism (SNP) arrays, DNA methylation arrays, RNA sequencing (RNA-seq) analyses and gene expression arrays. Finally, we have performed clinical studies aimed at translating the observed molecular alterations into clinical applications for CLL patients.

Mutational signatures in CLL subtypes

We studied pre-treatment tumour and matched non-tumour samples from 506 patients (452 CLL and 54 MBL): 317 (62%) were IGHV-mutated (IGHV-MUT), 179 (35%) IGHV-unmutated (IGHV-UNMUT), and 10 (2%) undetermined (Extended Data Table 1 and Supplementary Table 1). We performed WGS of 150 tumour/normal pairs, and WES of 440 cases (including 84 with both WGS and WES data). Somatic mutations analysed using the Sidrón pipeline[10] revealed the presence of 359,456 substitutions and small indels in

WGS analyses (240–5,416 per tumour), and an average mutation burden of 0.87 mutations per megabase (Mb) (Extended Data Fig. 1 and Supplementary Table 2). CLL and MBL samples had a similar mutation burden (0.87 versus 0.89 mutations Mb^{-1}, respectively, $P = 0.8$), and were considered together for WGS analysis. The number of somatic substitutions (excluding *IG* loci) was higher in IGHV-MUT tumours than in IGHV-UNMUT cases (2,847 versus 1,975, $P < 3 \times 10^{-8}$) (Extended Data Fig. 1). Three main mutational signatures were identified (Extended Data Fig. 1): an age-related signature involving C-to-T transitions at CpG sites; signature 2, characterized by T:A > G:C transversions; and an activation-induced cytidine deaminase (AID) signature[20]. This latter pattern was only detected on *IG* loci, although we also confirmed AID-induced mutations in some off-target genes highly expressed in the germinal centre[21,22]. Signature 2 was almost exclusively present in IGHV-MUT tumours, and its presence clearly separated IGHV-MUT from IGHV-UNMUT tumours (Extended Data Fig. 1).

Landscape of somatic mutations

We combined somatic mutations from the 506 tumour/normal pairs detected by either WGS or WES (excluding *IG* genes), resulting in a total of 13,631 somatic mutations affecting protein-coding genes

(average 26.9 per tumour) and 951 copy number alterations (CNAs) (average 1.9) (Fig. 1 and Supplementary Table 3). We identified 36 genes (tier 1) as recurrently mutated in CLL (false discovery rate (FDR) < 10%), and 23 additional genes (tier 2) were significantly mutated in one subgroup (IGHV-MUT or IGHV-UNMUT), had recurrent or truncating mutations, or had driver mutations described in other malignancies (Extended Data Table 2). Two genes (*BTG2* and *DTX1*) were excluded as they are known targets of the SHM machinery[21]. The remaining genes included most of the drivers previously described by different WES studies[9,11,13]. The most frequently mutated gene in CLL was *NOTCH1* (57 cases, 12.6%), followed by *ATM* (11%), *SF3B1* (8.6%), *BIRC3* (8.8%), *CHD2* (6%), *TP53* (5.3%) and *MYD88* (4%). Furthermore, we identified 12 novel genes recurrently mutated in CLL and not previously linked to this disease, including *ZNF292*, *ARID1A*, ZMYM3 and *PTPN11*. Most CLL driver genes were preferentially mutated in IGHV-UNMUT tumours and had subclonal mutations[11] (Supplementary Fig. 1). Notably, a similar frequency of mutated drivers was found in CLL and MBL cases of similar IGHV gene SHM status (Extended Data Table 2).

We also identified some genes (tier 3) that probably contain driver mutations but were found in three or less CLL patients. This is the case of activating mutations in the oncogenes *KRAS* and *NRAS*, truncating
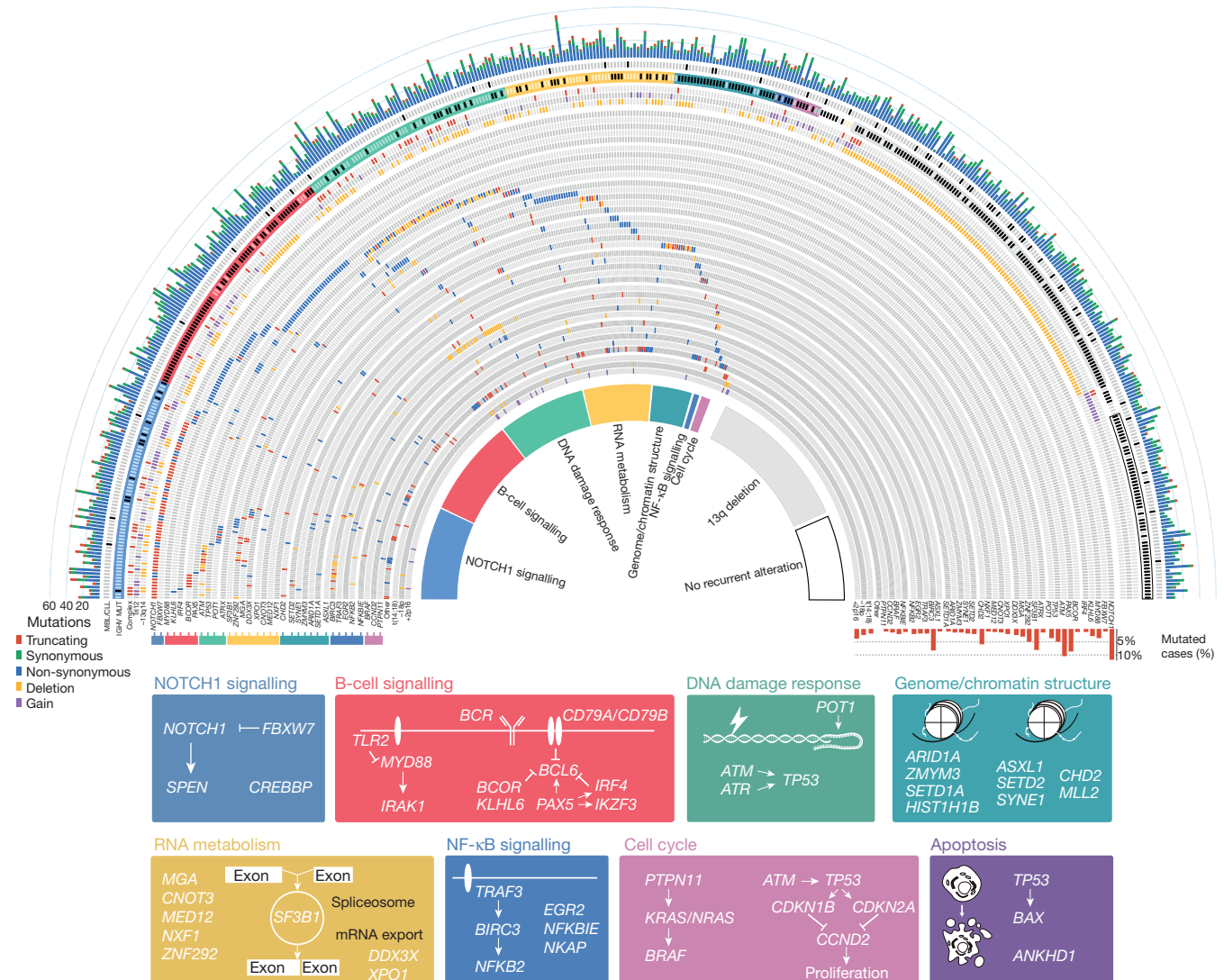

Figure 1 | Significantly mutated genes and pathways. The outer layer represents the number of truncating, non-synonymous and synonymous mutations for 506 CLL (grey) and MBL (black) cases. Clinical classification as well as IGHV-status is shown on the two outermost layers. Inner layers show the most recurrently mutated genes grouped according to the biological pathways involved (bottom). The percentage of cases with mutations in each gene is shown on the right.

mutations in the tumour suppressors *CDKN1B* and *CDKN2A*, and recurrent mutations in the transcription factor *IKZF3*. Mutations in components of the BCR and Toll-like receptor pathway were exclusively present in IGHV-MUT tumours. They included those in *MYD88*, *CD79A*, *CD79B*, *TLR2* and *IRAK1*, detected in 22 of the 278 IGHV-MUT cases, but in none of the 166 IGHV-UNMUT CLL patients ($P = 4.1 \times 10^{-5}$), confirming the importance of the BCR and Toll-like receptor pathways both in CLL pathobiology and as therapeutic targets[23]. Collectively, eight main pathways are frequently altered in CLL, including BCR signalling, cell cycle regulation, apoptosis, DNA damage response, chromatin remodelling, NF-κB signalling, NOTCH1 signalling, and RNA metabolism (Fig. 1).

DNA structural alterations

Analysis of structural variants confirmed the presence of known CNAs such as loss of 13q14, 11q22-q23, 17p, 6q15-q21 and trisomy 12 (Extended Data Fig. 2 and Supplementary Table 4). In addition, we identified novel candidate CLL driver genes in regions of recurrent chromosomal alterations (Fig. 1). They included deletions involving *ZNF292* at 6q15 (2.4%), deletions of 2q37 encompassing *SP140* and *SP110*, loss of 3p21 (2%) affecting *SMARCC1* and *SETD2*, and loss of 10q24 (1.8%) involving *NFKB2* (Supplementary Fig. 2).

Unlike other B-cell malignancies, translocations involving *IG* genes were uncommon in CLL with the exception of *BCL2* rearrangements (10 cases). They occurred exclusively in IGHV-MUT cases, and resulted in overexpression of *BCL2* and recruitment of the SHM machinery (Extended Data Fig. 3). Analysis of WGS data using SMUFIN[24] also revealed the presence of 147 interchromosomal translocations in 43 out of 148 cases (Supplementary Table 5). Recurrent translocations involving chromosome 13q14 with different chromosomal partners and associated with deletion or disruption of the microRNA cluster miR-15a/miR-16 were identified in nine cases ($P < 10^{-8}$). We also detected 15 non-recurrent chromosomal translocations, one of them involving the *IG* locus (*IGH-CBFA2T3*), and 14 predicted to originate in chimaeric genes, five of which could be confirmed by RNA-seq (Supplementary Table 5).

Complex rearrangements (chromothripsis/chromoplexy)[25,26] were identified in 15 out of 452 CLL cases (Extended Data Fig. 3), being more frequent in IGHV-UNMUT than in IGHV-MUT tumours (6% versus 1.8%, $P < 0.05$). Although these complex alterations did not result in any recurrent rearrangement, we observed involvement of chromosome 13 in 4 out of 15 tumours, resulting in *mir-15a/mir-16* loss. Similar to previous studies[27], mutations in *TP53* were more frequent in tumours with chromothripsis (26% versus 4.6%, $P < 0.006$). Furthermore, *SETD2* inactivation was more frequent in CLL cases

with chromothripsis than in non-chromothriptic cases (26% versus 1.4%, $P < 2 \times 10^{-4}$).

This analysis revealed significant relationships between several alterations, including co-occurrence of *NOTCH1* mutations and chromosome 12 trisomy[28], trisomy 12 with trisomy 18 ($q < 0.01$), and the mutually exclusive pattern of 13q14 deletion and trisomy 12 ($q < 0.01$). We also observed a higher co-occurrence of mutations in *NOTCH1* with those in *MGA* ($q < 0.01$), *BCOR* ($q < 0.01$) and *BIRC3* ($q < 0.05$), or gain of 2p16 with loss of 18p ($q < 0.01$), among others (Supplementary Fig. 3).

Mutations in non-coding regions

The presence of functional mutations outside of protein-coding regions remains an open question in cancer research[29]. We observed in one CLL case a previously described mutation in the *TERT* promoter (C228T)[29]. Eight mutations in *mir-142* were identified in five cases (Supplementary Fig. 4), with seven of them within AID target consensus (WRCY or WA), reinforcing it as a target of the SHM[30]. We also identified 88 mutations in non-coding regions present in at least two WGS cases (Supplementary Table 6). Most of them were located either within hypermutated late-replication regions[31], or within the 5′-region of *BACH2*, *BCL6*, *BTG2*, *CXCR4* and *TCL1A*, genes known to undergo SHM during the germinal centre reaction[21,22]. Most mutations were within the AID target sequence (WRCY), probably reflecting the passage of the respective progenitor cells through the germinal centre.

Notably, the most frequent recurrent non-coding mutation was detected in the 3′ UTR of *NOTCH1* (chr9: 139390152T > C), present in 4 of the 150 cases with WGS data (Fig. 2a). Sequencing of this region in 356 cases with only WES data revealed seven additional tumours with the same mutation, and two cases with a mutation seven or nine bases downstream of the original one. RNA-seq from six of these 3′ UTR *NOTCH1*-mutated tumours confirmed the presence of a novel splicing event within the last exon of *NOTCH1* (Fig. 2a), which was absent in 290 tumours without these mutations (Extended Data Fig. 4). This splicing event occurred preferentially between a cryptic donor site located in the coding region of the last exon of *NOTCH1* and a newly created acceptor site in the 3′ UTR, resulting in a deletion that includes the last 158 coding bases. Nevertheless, some splicing events occurred between the canonical donor site on exon 33 and the newly created acceptor site in the 3′ UTR of exon 34 (Fig. 2a). Reverse transcription PCR (RT–PCR) analysis confirmed the presence of this aberrant splicing only in cases with mutations in the 3′ UTR (Fig. 2b). This within-exon splicing is predicted to remove a PEST domain of NOTCH1 and to increase protein stability, as in the previously
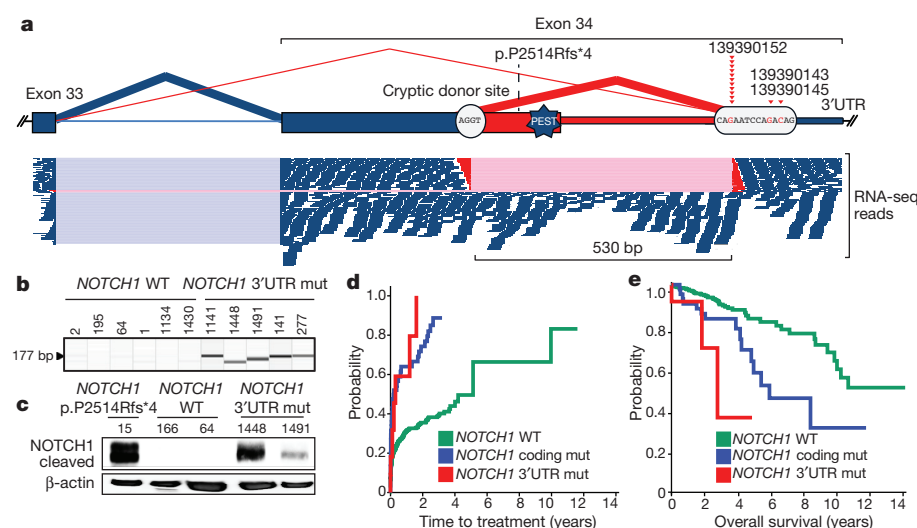

Figure 2 | Activating mutations in the 3′ UTR non-coding region of *NOTCH1*. a, Mutant bases are shown in red and the number of cases denoted by arrowheads. Aberrant spliced reads detected by RNA-seq (red) are shown below. **b**, RT–PCR amplification shows the expected 177-base-pair (bp) band in tumours with the recurrent 139390152T > C mutation, and a smaller one in cases with the 139390145 and 139390143 mutations. WT, wild type. **c**, Western blot analysis showing the accumulation of a lower molecular mass NOTCH1 protein in CLL cells with the p.P2514Rfs*4 or the 3′ UTR mutation. β-actin was used as loading control. **d**, **e**, Kaplan–Meier plot of time-to-treatment (**d**) or overall survival (**e**) of CLL patients grouped on the basis of mutations in the 3′ UTR of *NOTCH1*, the presence of *NOTCH1* coding mutations, or *NOTCH1* wild type.

described p.P2514Rfs*4 *NOTCH1* mutation[10]. Western blot analysis confirmed the presence of a smaller molecular mass band in 3′-UTR- and p.P2514Rfs*4-mutated cells, which was absent in cells without mutations in *NOTCH1* (Fig. 2c). Immunohistochemical analysis showed a strong NOTCH1 nuclear signal in tumour cells from patients with 3′ UTR or p.P2514Rfs*4 mutations (Extended Data Fig. 4). All cases with mutations in the 3′ UTR of *NOTCH1* belonged to the IGHV-UNMUT subgroup, accounting for up to 6.7% (12 out of 179) of all IGHV-UNMUT cases. Patients with 3′ UTR *NOTCH1* mutations had features of adverse prognosis (Extended Data Fig. 4) and behaved similarly to patients with coding mutations in *NOTCH1* in terms of the time to first treatment (TTT) and overall survival (Fig. 2d, e).

We further explored the presence of genome regions with high mutational density and found 24 loci enriched in somatic mutations (Fig. 3a). Most of them correspond either to recurrently mutated genes in CLL or to known targets of the SHM process. However, we identified a densely mutated cluster in a small intergenic region of chromosome 9p13, in which 17 different tumours had somatic mutations (Fig. 3b). This region is enriched for both lymphocyte-specific transcription factor binding sites and histone marks related to enhancer elements only in a lymphoblastoid B-cell line (Supplementary Fig. 5). DNase-seq and chromatin immunoprecipitation sequencing (ChIP-seq) analysis in normal B cells and CLL cases revealed that the region contains an active enhancer characterized by a DNase I hypersensitive site and nucleosomes containing histone 3 Lys4 methylation (H3K4me1) and H3K27 acetylation (H3K27ac) (Fig. 3b and Supplementary Fig. 5). Chromosome conformation capture sequencing (4C-seq) analysis[32] in tumour cells from two CLL patients revealed that this potential enhancer shows high three-dimensional contact frequencies extending towards the telomere up to the *PAX5* locus, located 330 kilobases (kb) away (Fig. 3c and Supplementary Fig. 5). Expression analysis of 15 genes located within 1 Mb of this element revealed that the only gene showing a significant

expression difference correlated with the presence of mutations within the putative enhancer region was indeed *PAX5* (average expression 87 versus 131, $P = 1.9 \times 10^{-4}$) (Extended Data Fig. 5). *PAX5* encodes a transcription factor that has an essential role in B-cell differentiation[33] and, based on the evidence provided above, is the most likely target of the identified enhancer region. CRISPR/Cas9-based genome editing of this region allowed us to demonstrate that either the introduction of a specific point mutation, or the deletion of this putative enhancer in a lymphoblastoid B-cell line or in RAMOS cells, resulted in a 40% reduction in the expression of *PAX5* (Extended Data Fig. 6).

Sequencing of this region in all CLL cases with WES data identified 25 new cases with somatic mutations. We also found somatic mutations in this enhancer in diffuse large B-cell lymphomas (29%, 26 out of 89), follicular lymphomas (23%, 20 out of 86) and mantle-cell lymphomas (5%, 3 out of 66) (Supplementary Table 7). Interestingly, 84% of CLL cases with mutations in this enhancer belong to the IGHV-MUT subgroup, accounting for up to 13% of IGHV-MUT CLL cases. Mutations in the *PAX5* enhancer were the only recurrent alteration observed in 7 cases, while in 11 tumours this alteration was only combined with 13q14 deletion, raising the possibility that *PAX5* enhancer mutations might constitute driver events contributing to the development of these tumours.

Integrative analysis

We then integrated the standard genetic classification of CLL with a recent patient categorization in three subgroups based on a DNA methylation signature of naive and memory B cells[17,34] (Supplementary Table 1). The three epigenetic subgroups showed a distinct distribution of genetic changes, IGHV gene repertoire and stereotyped B-cell receptors (Extended Data Fig. 7). The intermediate group had moderate IGHV mutation levels, an intermediate contribution of signature 2 mutations, higher frequencies of *SF3B1* and *MYD88* mutations, biased usage of the IGHV-3-21 and IGHV-1-18 genes
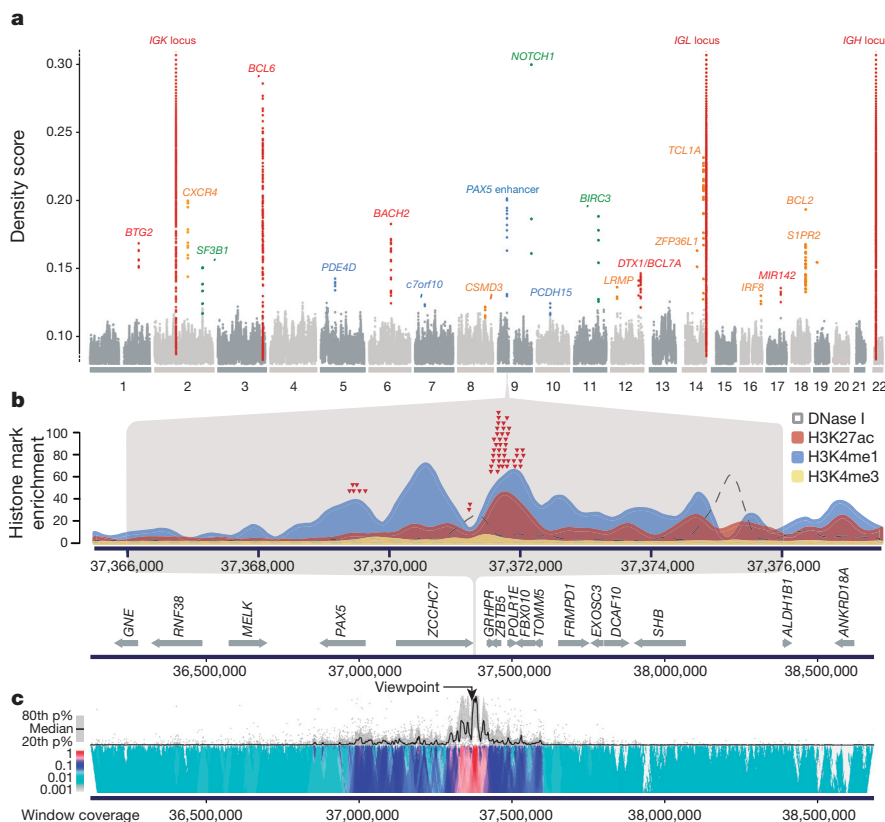

Figure 3 | Identification of somatic mutations in a *PAX5* enhancer. a, Regions with a high density of somatic mutations in 150 WGS analyses. Regions correspond to recurrently mutated genes (green), targets of SHM (red/orange), and other regions (blue). **b**, Detailed view of a 9p13 region showing the accumulation of somatic mutations (arrowheads) in CLL tumours as well as DNase I hypersensitivity and histone H3K27ac, H3K4me1 and H3K4me3 enrichment from CLL tumour 110. **c**, 4C-seq analysis in CLL cells showing the interaction frequencies of the enhancer with the surrounding regions. p%, percentile.

and increased frequency of stereotyped subset #2. These results support the hypothesis that this group has a distinct genetic and epigenetic makeup[17,34–36]. We also found a highly significant correlation ($r = 0.64$, $P < 0.001$) between the number of WGS mutations per case and the number of CpGs showing differential methylation as compared to naive B cells (Extended Data Fig. 7). Similarly, the proportion of signature 2 mutations was also correlated with differential methylation in IGHV-MUT cases.

MBL cases were indistinguishable at the genomic, transcriptomic and epigenomic level from CLL cases assigned to the same IGHV subgroup (Extended Data Fig. 7 and Extended Data Table 2), in accordance with the overlapping biological features of both processes. Notably, the burden of driver alterations was significantly lower in patients with MBL than with CLL (1.2 versus 1.7, for IGHV-MUT cases, $P = 8 \times 10^{-4}$), consistent with a model in which MBL/CLL evolution is accomplished by the progressive accumulation of driver alterations.

Clinical implications

Our data support the hypothesis that the observed genomic differences between the two major molecular subgroups of CLL might be in part responsible for their different outcome. The average number of driver mutations in IGHV-UNMUT tumours was higher than in IGHV-MUT cases (3.5 versus 1.7, $P < 10^{-19}$), despite the 44% higher mutational burden of IGHV-MUT tumours. We found that 88% of cases had at least one driver mutation, with almost all IGHV-UNMUT tumours containing at least one driver alteration, while a smaller fraction was found in the IGHV-MUT subgroup (96% versus 83%, $P < 5 \times 10^{-5}$).

We evaluated the influence of the presence of each alteration on the TTT and overall survival from the time of sampling. The mutation of several drivers and CNAs was significantly correlated with an adverse prognosis, in some cases independently from Binet stage and IGHV mutational status (Fig. 4a, Extended Data Fig. 8 and Supplementary Table 8). We confirmed the independent prognostic value of known gene mutations (SF3B1 and TP53), and identified novel independent

prognostic drivers for both shorter TTT (BRAF, ZMYM3, IRF4, NFKB2, 20p deletion, and 2p16 and 5q34 gains), and overall survival (ASXL1, POT1 and 14q24 deletion). Remarkably, the accumulative number of drivers (0 to ≥4) per tumour had a progressively worse effect on outcome that could discriminate patient subsets differing by more than 10 years in the median TTT, independently of IGHV status and Binet stage. They also showed prognostic value for overall survival, although not independent in the multivariate analysis (Fig. 4b, c). Finally, we examined the potential druggability of the alterations in genes and pathways identified in CLL patients[37], finding candidate drugs for 19 of the 59 driver genes in 42% of the CLL cases (190 out of 452) (Supplementary Fig. 6 and Supplementary Tables 9 and 10).

Discussion

In this work, we have provided a comprehensive and integrated molecular characterization of CLL. We have also unveiled new biological aspects of this disease and identified novel driver genes presumably implicated in its pathogenesis. The large number of different genomic alterations found in our cohort illustrates the enormous biological heterogeneity of CLL. Notably, the use of WGS has allowed us to identify recurrent mutations in non-coding regions, including the 3′ UTR of NOTCH1 and a PAX5 enhancer, resulting in marked alterations in the activity of these transcription factors of well-known importance in leukaemia and other malignancies[38,39]. Previous studies have shown the effect of NOTCH1 mutations in CLL prognosis[10,40]. However, these studies may seriously underestimate the true incidence of NOTCH1 deregulation in CLL, given our finding that about 20% of NOTCH1-mutated tumours contain mutations in the 3′ non-coding region. These findings emphasize the value of large genome-wide studies to discover new molecular alterations that may have a profound effect on cancer development and progression.

The evaluation of putative associations between these molecular alterations and the clinicopathological features of our cohort of CLL patients has been challenging owing to the low frequency of many significantly mutated genes. Patients in which no recurrent alterations
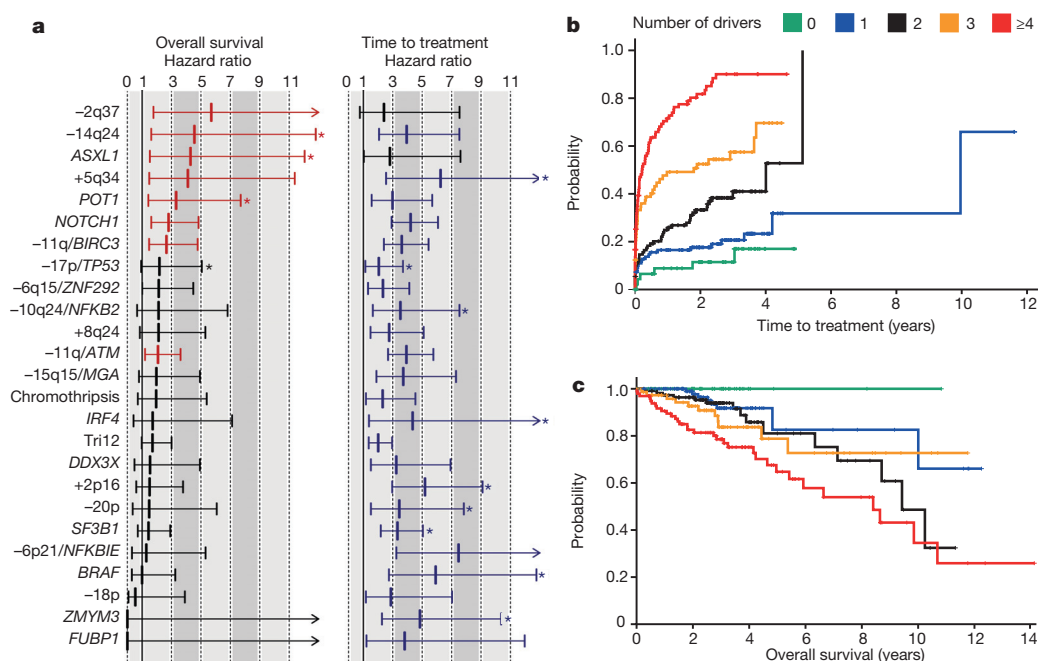

Figure 4 | Prognostic effects of individual alterations and number of drivers. a, Effect on overall survival (left) and time-to-treatment (right) for each genomic alteration. Labels including genes and chromosomal regions represent combined analysis of mutations and copy number alterations. Hazard ratios and 95% confidence intervals are shown. Alterations conferring statistically significant (adjusted $P < 0.05$) hazard ratios are shown in colour (red for overall survival and blue for TTT), and those in which the effect was independent of Binet stage and IGHV-status are labelled with an asterisk. **b, c**, Kaplan–Meier plots of TTT (**b**) or overall survival (**c**) of CLL patients grouped by the number of driver mutations identified.

were found had the best prognosis and near normal overall survival, suggesting that this study has uncovered most driver alterations involved in CLL evolution, opening new avenues to explore the clinical impact of the heterogeneous molecular composition of the disease in independent cohorts. Hopefully, this work will finally result in new opportunities for improving the clinical management and personalized treatment of CLL patients.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Gaidano, G., Foa, R. & Dalla-Favera, R. Molecular pathogenesis of chronic lymphocytic leukemia. *J. Clin. Invest.* **122,** 3432–3438 (2012).
2. Zenz, T., Mertens, D., Kuppers, R., Dohner, H. & Stilgenbauer, S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature Rev. Cancer* **10,** 37–50 (2010).
3. Pekarsky, Y., Zanesi, N. & Croce, C. M. Molecular basis of CLL. *Semin. Cancer Biol.* **20,** 370–376 (2010).
4. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V$_H$ genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94,** 1848–1854 (1999).
5. Damle, R. N. et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94,** 1840–1847 (1999).
6. Crespo, M. et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N. Engl. J. Med.* **348,** 1764–1775 (2003).
7. Malek, S. N. The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. *Oncogene* **32,** 2805–2817 (2013).
8. Döhner, H. et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343,** 1910–1916 (2000).
9. Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia. *Nature Genet.* **44,** 47–52 (2011).
10. Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475,** 101–105 (2011).
11. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152,** 714–726 (2013).
12. Fabbri, G. et al. Analysis of the chronic lymphocytic leukemia coding genome: role of *NOTCH1* mutational activation. *J. Exp. Med.* **208,** 1389–1401 (2011).
13. Ramsay, A. J. et al. *POT1* mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nature Genet.* **45,** 526–530 (2013).
14. Damm, F. et al. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov.* **4,** 1088–1101 (2014).
15. Rossi, D. et al. Disruption of *BIRC3* associates with fludarabine chemorefractoriness in *TP53* wild-type chronic lymphocytic leukemia. *Blood* **119,** 2854–2862 (2012).
16. Ferreira, P. G. et al. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res.* **24,** 212–226 (2014).
17. Kulis, M. et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nature Genet.* **44,** 1236–1242 (2012).
18. Oakes, C. C. et al. Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov.* **4,** 348–361 (2014).
19. Hudson, T. J. et al. International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
20. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
21. Khodabakhshi, A. H. et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget* **3,** 1308–1319 (2012).
22. Pasqualucci, L. et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412,** 341–346 (2001).
23. Byrd, J. C. et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* **369,** 32–42 (2013).
24. Moncunill, V. et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nature Biotechnol.* **32,** 1106–1112 (2014).
25. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).
26. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153,** 666–677 (2013).
27. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148,** 59–71 (2012).
28. Balatti, V. et al. *NOTCH1* mutations in CLL associated with trisomy 12. *Blood* **119,** 329–331 (2012).
29. Huang, F. W. et al. Highly recurrent *TERT* promoter mutations in human melanoma. *Science* **339,** 957–959 (2013).
30. Yamane, A. et al. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature Immunol.* **12,** 62–69 (2011).
31. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
32. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nature Methods* **4,** 895–901 (2007).
33. Revilla-i-Domingo, R. et al. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J.* **31,** 3130–3146 (2012).
34. Queirós, A. C. et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* **29,** 598–605 (2015).
35. Strefford, J. C. et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of *SF3B1* and subset #2. *Leukemia* **27,** 2196–2199 (2013).
36. Agathangelidis, A. et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood* **119,** 4467–4475 (2012).
37. Rubio-Perez, C. et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* **27,** 382–396 (2015).
38. Lobry, C., Oh, P. & Aifantis, I. Oncogenic and tumor suppressor functions of Notch in cancer: it's NOTCH what you think. *J. Exp. Med.* **208,** 1931–1935 (2011).
39. O'Brien, P., Morin, P. Jr, Ouellette, R. J. & Robichaud, G. A. The Pax-5 gene: a pluripotent regulator of B-cell differentiation and cancer disease. *Cancer Res.* **71,** 7345–7350 (2011).
40. Villamor, N. et al. *NOTCH1* mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia* **27,** 1100–1106 (2013).

Supplementary Information is available in the online version of the paper.

Author Contributions The Chronic Lymphocytic Leukaemia Genome consortium contributed to this study as part of the International Cancer Genome Consortium. Investigator contributions are as follows: T.B., J.D., A.L.-G., A.R.P., M.G. and J.M.H.-R. contributed to sample collection and clinical annotation; M.R., N.V., E.Ca., E.Co., J.M.H.-R. and M.G. were the pathologists who reviewed and confirmed the diagnoses; P.N., C.M.R.-C. and M.A. prepared and supervised the bioethical requirements; M.P., A.E. and C.R. processed samples and performed validation analysis; M.G., I.G. and D.A.P. were responsible for generating libraries, performing exome capture and sequencing; S.B., D.To., M.M., S.G., I.S., G.C., D.M.-G., A.C., X.E. and D.Cos. analysed copy number alterations and structural variants; X.S.P., R.V.-M., J.G.-A. and V.Q. developed the bioinformatic pipeline for analysis of somatic mutations and performed functional data integration; D.Col., M.L.-G. and B.G. were responsible for downstream validation analysis and functional studies; A.N. and K.S. analysed *IG* gene rearrangements and stereotypes; J.I.M.-S., A.C.Q., G.C., R.B., R.G., N.R., H.G.S. and P.J. performed epigenetic and transcriptomic analysis and 4C-seq experiments; L.B. and A.A.F. performed enhancer analysis and CRISPR experiments; N.V., T.B., A.L.-G. and E.Ca. performed clinical and biological studies; J.L.G., R.R., M.O., D.G.P. and A.V. were in charge of bioinformatics data management; N.L.-B., C.R.-P. and D.Ta. contributed to pathway analysis and *in silico* prescription. X.S.P., C.L.-O. and E.Ca. directed the research, analysed the data and wrote the manuscript.

Author Information Sequencing, expression and genotyping array data have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00000000092. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.L.-O. (clo@uniovi.es) or E.C. (ecampo@clinic.ub.es).

METHODS

Patients. The clinical and biological characteristics of the 506 patients are shown in Extended Data Table 1. Among these patients, 452 were diagnosed with CLL and 54 with MBL. Cases were defined as IGHV-MUT when the identity of immunoglobulin genes was less than 98%. The tumour samples were obtained before administration of any treatment. All patients gave informed consent for their participation in the study following the International Cancer Genome Consortium (ICGC) guidelines and the ICGC Ethics and Policy committee[19].

Collection and preparation of samples. Tumour samples were obtained from fresh or cryopreserved mononuclear cells. To purify the CLL or MBL fraction, samples were incubated with a cocktail of magnetically labelled antibodies directed against T cells, natural killer cells, monocytes and granulocytes (CD2, CD3, CD11b, CD14, CD15 and CD56), adjusted to the percentage of each contaminating population (AutoMACS, Miltenyi Biotec). The degree of contamination by non-CLL cells in the CLL fraction was assessed by immunophenotype and flow cytometry. DNA was extracted from purified samples by using a Qiagen kit, and the quality of purified DNA was assessed by SYBR-green staining on agarose gels and quantified using a Nanodrop ND-100 spectrophotometer. The tumour DNA and RNA samples for further genomic analysis contained ≥95% neoplastic cells and the contamination by neoplastic cells in normal DNA was <2%.

WGS, WES and RNA-seq. For WGS, 2 μg of genomic DNA from each sample was used for the construction of two short-insert paired-end sequencing libraries. One library was prepared using a standard TruSeqDNA Sample Preparation Kit v2 (Illumina Inc.) with some modifications. In short, following the fragmentation (CovarisE220) the libraries were size-selected on the agarose gel and processed through end-repair, adenylation and indexed adaptor ligation. The gel eluate was directly amplified by 10 PCR cycles. The second library was prepared following the same protocol as above, however, it included a heating step to 72 °C before adaptor ligation and was suddenly cooled down to 4 °C. This resulted in a biased proportion of high GC content reads and counterbalanced some of Illumina's PCR sample preparation methods' GC-bias, thus improving coverage of increased GC-content regions of the genome. Both types of libraries were sequenced in paired-end mode on Illumina GAIIx (2 × 151 bp) using Sequencing kit v4 or Illumina HiSeq2000 (2x101 bp) using TruSeq SBS Kit v3 (Illumina Inc.).

For other samples (Supplementary Table 1), the library preparation procedure was modified to remove the PCR step during short-insert paired-end library preparation. The TruSeq DNA Sample Preparation Kit v2 (Illumina Inc.) and the KAPA Library Preparation kit (Kapa Biosystems) were used. In brief, 2 μg of genomic DNA was sheared on a Covaris E220, size-selected and concentrated using AMPure XP beads (Agencourt, Beckman Coulter) to reach the fragment size of 220–480 bp. Fragmented DNA was end-repaired, adenylated and ligated to Illumina specific indexed paired-end adaptors. All libraries were quantified by Library Quantification Kit (Kapa Biosystems). Each library was sequenced using TruSeq SBS Kit v3-HS (Illumina Inc.), in paired-end mode, 2 × 101-bp, in three sequencing lanes of HiSeq2000 flowcell v3 (Illumina Inc.) according to standard Illumina operation procedures with minimal yield of 85 Gb for each sample. Primary data analysis was carried out with the standard Illumina software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ files.

For WES, 3 μg of genomic DNA from each sample were sheared and used for the construction of a paired-end sequencing library as described in the paired-end sequencing sample preparation protocol provided by Illumina[41]. Enrichment of exonic sequences was then performed for each library using either the Sure Select Human All Exon 50 Mb or All Exon+UTRs v4 kits (Supplementary Table 1) following the manufacturer's instructions (Agilent Technologies). Exon-enriched DNA was pulled down by magnetic beads coated with streptavidin (Invitrogen), followed by washing, elution and 18 additional cycles of amplification of the captured library. Enriched libraries were sequenced (2 × 76 bp) in one lane of an Illumina GAIIx sequencer or in two lanes of a HiSeq2000 when using pools of eight samples.

RNA was assayed for quantity and quality using Qubit RNA HS Assay (Life Technologies) and RNA 6000 Nano Assay on a Bioanalyzer 2100. RNA-seq libraries were prepared from total RNA using the TruSeq RNA Sample Prep Kit v2 (Illumina Inc.) with minor modifications. In brief, 0.5 μg of total RNA was used as the input material for poly-A-based messenger RNA enrichment with oligo-dT magnetic beads. Selected mRNA was fragmented (resulting RNA fragment size was 80–250 nucleotides, with the major peak at 130 nucleotides). After first and second strand cDNA synthesis the double-stranded complementary DNA was end-repaired, 3′ adenylated and the 3′ 'T' nucleotide of the adaptor was used for the Illumina indexed adapters ligation. The ligation product was enriched by 10 cycles of PCR. Each library was sequenced using TruSeq SBS Kit v3-HS, in paired-end mode with a read length of 2 × 76 bp. We generated more than 20 million paired-end reads for each sample in a fraction of a sequencing lane on HiSeq2000 (Illumina Inc.) following the manufacturer's protocol. Image

analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ sequence files.

Read mapping and processing. For WGS and WES, reads from each library were mapped to the human reference genome (GRCh37) using BWA[42] with the same option, and a BAM file was generated using SAMtools[43]. Reads from the same paired-end libraries were merged, and optical or PCR duplicates were flagged using Picard (http://picard.sourceforge.net/index.shtml). For the identification of somatic substitutions and indels, we used the Sidrón algorithm[9,44]. This algorithm was adapted to identify subclonal mutations in which the mutant allele fraction is low, but supported by at least three reads. Visual inspection of recurrent mutational hotspots allowed the inclusion of some somatic mutations that were originally discarded owing to the presence of an excess of mutant reads in the non-tumour sample, or owing to low coverage, especially in the case of *NOTCH1*, in which a high GC content on exon 34 usually resulted in very low coverage by WES. In samples in which *NOTCH1* coverage was too low to make a call, mutations were analysed by Sanger sequencing. A comparison of mutation calls by Sidrón and by Sanger sequencing of some of the most frequently mutated genes in CLL (*SF3B1*, *TP53*, *MYD88*) revealed more than 97% specificity and at least 90% sensitivity. Mutational signatures were extracted using the WTSI Mutational Signature Framework[45]. To estimate the presence of subclonal mutations in recurrently mutated genes, the fraction of reads supporting a mutant allele was calculated for those mutations in which the depth of coverage was at least 20 reads. Flow cytometry analysis confirmed that the percentage of tumour cells was at least 98%. A case was considered as having a clonal mutation when at least 80% of cells were estimated to contain the mutation, and the mutant allelic fraction was within the 95% confidence interval.

Analysis of CNAs and structural variants. For the identification of CNAs, tumour and normal DNA from 505 CLL patients were analysed using Affymetrix SNP6.0 microarrays (Affymetrix) as previously described[46]. SNP array experiments were carried out at CeGen (http://www.cegen.org). Additionally, for 230 cases array-comparative genomic hybridization was performed in SurePrint G3 Human aCGH Microarray 1M (Agilent Technologies). Array-comparative genomic hybridizations were performed at qGenomics (http://www.qgenomics.com). Nexus 6.0 Discovery Edition software (Biodiscovery) was used for global analysis and visualization. Copy number neutral loss of heterozygosity was considered when the size of alteration was larger than 5 Mb. Acquired copy number neutral loss of heterozygosity was observed in 28 regions, 16 of them affecting known driver genes that already contained mutations, resulting in homozygous deletion of *mir-15a/mir-16* at 13q14, or inactivation of *ATM* and *TP53* (Supplementary Table 4). According to the literature, the presence of chromothripsis was considered when at least seven switches between two or more copy number states were detected on an individual chromosome in which LOH was retained, and chromoplexy was defined when at least three chained chromosomal rearrangements were detected in a tumour[27,47]. In one case in which genotyping data were not available, we used exome2cnv[48] to identify CNAs from WES data.

For the identification of breakpoints in WGS derived from structural variants, we used SMUFIN[24], a program that directly compares sequence reads from normal and tumour samples, to identify chromosomal breakpoints corresponding to large structural variants at base-pair resolution. We analysed 150 tumour/normal whole-genome pairs setting the cross-sample contamination filter to 5%. Two WGS tumours (019 and 029) showed an abnormal number of breakpoints owing to the presence of sequence lanes with high error rates that interfere with SMUFIN and were not considered for this analysis. All predicted breakpoints that were not confirmed through the BAM file after manual inspection were systematically discarded. A total of 48 out of 53 (91%) selected predicted breakpoints could be verified using PCR amplification followed by Sanger sequencing (Supplementary Table 5). This verification rate is similar to the one observed in our initial description of the method[24]. In addition, custom scripts were used to identify potential translocations involving immunoglobulin genes either in WGS or WES. This resulted in the identification of ten cases (5 WGS and 5 WES) containing putative translocations with the *BCL2* locus (nine with the t(14;18)(q32;q21), and one with the t(2;18)(p11;q21) translocation), all of which were confirmed by either Fluorescence *in situ* hybridization (FISH), cytogenetics or PCR (Extended Data Fig. 3).

G-banding and FISH analysis. Conventional cytogenetics was performed on Giemsa-banded chromosomes (G-banding) obtained after a 72-h culture and stimulation with tetradecanoyl-phorbol-acetate. At least 20 G-banded metaphases per sample were analysed. Results were described according to the International System for Human Cytogenetic Nomenclature. FISH analyses on fixed cells were performed using probes that interrogated for 11q23/*ATM*, 13q14.3 and 17p13/*TP53* deletions and trisomy 12 (Abbott Molecular). Two hundred nuclei were examined for each probe. LSI *IGH/BCL2* dual colour fusion for the t(14;18)(q32;q21) (Abbot Molecular) was used to confirm *BCL2*

rearrangements detected by WGS and WES. Additionally, in case 853, whole chromosomal paintings of chromosomes 8, 11 and X were performed to determine the complex karyotype (with four derivative chromosomes), and rearrangements predicted by SMUFIN algorithm.

Analysis of DNA methylation. DNA methylation was analysed using the 450k Human Methylation Array (Illumina). We used the EZ DNA Methylation Kit (Zymo Research) for bisulphite conversion of 500 ng of genomic DNA, and the Infinium methylation assay was carried out as described by the manufacturer[49,50]. These array experiments were performed at CeGen (http://www.cegen.org). Data from the 450k Human Methylation Array were analysed in R using the minfi package (version: 1.6.0)[51], available through the Bioconductor open source software, applying several custom filters. Unsupervised analyses were performed by principal component analysis and differential methylation between individual CLL/MBL samples and controls was detected using an absolute difference of 0.25.

Gene expression profiling. We studied the gene expression profiling of 468 cases using highly purified leukaemic CLL cells. Total RNA was extracted with the TRIzol reagent following the recommendations of the manufacturer (Invitrogen Life Technologies). RNA integrity was examined with the Agilent 2100 Bioanalyzer (Agilent Technologies) and only high-quality RNA samples were hybridized to Affymetrix Human Genome Array U219 array plates according to Affymetrix standard protocols. Summarized expression values were computed using the robust multichip average approach implemented in the Expression Console Software (Affymetrix Inc.).

RT–PCR. cDNA was synthesized from 500 ng of total RNA using High Capacity RNA-to-cDNA kit (Life Technologies) following the manufacturer's instructions. Amplification was performed using 50 ng of DNA using Qiagen Multiplex PCR Kit (Qiagen), and the reaction mix contained $1\times$ Qiagen Multiplex PCR Master Mix (12.5 µl), primer mix (0.4 µM of each primer) and RNase-free water for a total reaction volume of 25 µl. For *NOTCH1* within-intron splicing, primers used were: forward 5′-CCTAACAGGCAGGTGATGCT-3′ and reverse 5′-TACTCTCGCCTGTGGACAA-3′. PCR amplification was performed for *NOTCH1* 3′ UTR forward primer 5′-CCTAACAGGCAGGTGATGCT-3′ and reverse primer 5′-ATCTGGCCCCAGGTAGAAAC-3′, *PAX5* enhancer first region forward 5′-TAGATTGTGCCGAATGCTGA-3′ and primer 5′-ACAAGCTCTCCTCCCAGGAA-3′, and *PAX5* enhancer second region forward primer 5′-AGGATGAGAACGGGCAAAC-3′ and reverse primer 5′-GGAGCTTCCAGCTGAACTGA-3′. All PCR products were run on a capillary electrophoresis gel (QIAxcel Advanced System, Qiagen) with the QIAxcel DNA screening kit (Qiagen).

Western blot analysis. For western blot analysis, tumour cells were lysed for 30 min in Triton buffer (1% Triton X-100, 50 mM Tris–HCl, pH 7.6, 150 mM NaCl, 1 mM EDTA) supplemented with protease and phosphatase inhibitors (1 mM PMSF, 2 mM sodium pyrophosphate, 2 mM sodium β-glycerophosphate, 1 mM NaF, 1 mM sodium orthovanadate, 10 µg ml^{-1} leupeptin and 10 µg ml^{-1} aprotinin). Lysates were cleared by centrifugation at 15,000g at 4 °C for 15 min, and protein concentrations determined using the Bradford method. Thirty micrograms of protein was separated by SDS–PAGE and transferred onto Immobilon-P membranes. Membranes were blocked with 2.5% phospho-Blocker (Cell Biolabs) in TBS-Tween 20. For protein immunodetection, the specific primary antibodies were used: anti-cleaved NOTCH1 (Val1744) (D3B8; Cell Signaling Technology) and β-actin (Sigma). Anti-rabbit and anti-mouse horseradish peroxidase-labelled IgG (Sigma) were used as secondary antibodies. Chemiluminescence was detected by using ECL substrate (Pierce) on a mini-LAS4000 Fujifilm device (GE Healthcare).

Immunohistochemical analysis. NOTCH1 immunohistochemical staining was performed on a Leica Bond system using formalin-fixed paraffin-embedded tissue sections[52]. Samples were pre-treated using heat-mediated antigen retrieval with EDTA buffer (pH 9.0), epitope retrieval solution 2 (HIER2) for 30 min. Then, sections were incubated with anti-cleaved NOTCH1 rabbit monoclonal antibody (clone D3B8, catalogue number 4147, Cell Signaling Technology) at a final concentration of 8.5 µg ml^{-1}, for 60 min at room temperature and detected using a horseradish peroxidase (HRP)-conjugated compact polymer system. DAB was used as the chromogen. The section was then counterstained with haematoxylin and mounted with DPX.

Sanger sequencing. PCR products were treated using ExoSap IT (USB Corporation) and sequenced with ABI Prism BigDye terminator v3.1 (Applied Biosystems) and 5 pmol of each primer. Sequencing reactions were run on an ABI-3730 automated sequencer (Applied Biosystems). All sequences were examined with the Mutation Surveyor DNA Variant Analysis Software (Softgenetics).

ChIP-seq and DNase-seq. ChIP-seq was performed in normal B-cell subpopulations and in cells (>90% tumour cell content) of a CLL patient with mutated IGHV, and DNase-seq only in the latter following standard protocols generated within the Blueprint Consortium. In brief, cells for ChIP-seq were fixed for

8–16 min in 1% formaldehyde at 4 °C, and chromatin was sonicated for 15 min with a Biorruptor (Diagenode). Chromatin fragments ranging from 50 to 500 bp were selected and immunoprecipitation was carried out with antibodies from Diagenode against H3K4me3 (pAb-003-050 lot:A5051-001P), H3K4me1 (pAb-194-050 lot:A1863-001P) and H3K27ac (pAb-196-050 lot: A1723-0041D) using approximately 500,000 cells per antibody. DNase I digestion was performed using 60 units of the enzyme (Sigma) and 2.5 million cells. ChIP-seq and DNase-seq libraries were constructed using the Kapa Hyper Prep Kit (Kapa Biosystems). For each experiment, from 25 to 50 million reads were sequenced with an Illumina HiSeq2000 sequencer. Detailed protocols can be obtained from the Blueprint Consortium (http://www.blueprint-epigenome.eu/index.cfm?p = 7BF8A4B6-F4FE-861A-2AD57A08D63D0B58).

4C-seq. 4C-seq template generation and amplification was performed as previously described[53,54]. In brief, 1×10^7 cells of two CLL patients were crosslinked with 2% formaldehyde (Merck), chromatin was digested with DpnII (New England Biolabs) followed by ligation with T4 ligase (Roche). Next, chromatin was decrosslinked, DNA was digested with Csp6I (NEB) and re-ligated. PCR amplification of viewpoint regions and their ligated fragments was performed using primers 5′-TGCCACACCTCCTTTTGATC-3′ and 5′-CCTTGTGGAAAGAGTCTCAC-3′ (*PAX5* putative enhancer, viewpoint fragment-end chr9:37,370,916-37,371,635) or 5′-CCGAGCTGGGGTAGCTGATC-3′ and 5′-TTGTGTCCAAAAGTTGTTTG-3′ (*PAX5* promoter, viewpoint fragment-end chr9:37,033,553-37,034,192). Samples were sequenced using a MiSeq instrument (Illumina) using 50-bp single-end reads, and adding 5% PhiX control DNA. Data analysis was performed using 4Cseqpipe version 0.7 (May 2012) (downloaded from http://compgenomics.weizmann.ac.il/tanay/). Before mapping of the interacting regions to the genome, reads that are a consequence of undigested templates or self-ligation of the viewpoint fragment were removed.

Deletion and mutation of human PAX5 enhancer in B-cell lines using CRISPR/Cas9. Human *PAX5* enhancer was deleted or mutated in RAMOS cells and in an Epstein–Barr virus (EBV)-transformed lymphoblastoid B-cell line using CRISPR/Cas9 genome editing. Guide RNAs (gRNAs) were designed using E-CRISP tool (http://www.e-crisp.org/E-CRISP/index.html)[55]. For the deletions, four gRNAs were designed flanking the *PAX5* enhancer, two at each side (L1/L2 and R1/R2) to be used in combinations (L1+R1, L1+R2, L2+R1, and L2+R2). In addition, two gRNAs were designed to target sites of mutations found in CLL (M1/M2). gRNAs sequences are: L1, 5′-GGGAACCAGGGCGTGGGAGC-3′; L2, 5′-GTGAGGCAGAAACACCACAG-3′; R1, 5′-GGCAGCATGCGGGCGTCATG-3′, R2, 5′-GCCAGGACCCTGCTCTCCCAA-3′; M1, 5′-GTGAAAATTTACTCATGCTG-3′; and M2, 5′-GGTGGTACTCAGAGGCTGGG-3′. The gRNA oligonucleotides were cloned in pL-CRISPR.EFS.GFP vector (Addgene plasmid 57818)[56], and lentiviral particles were produced on HEK293T cells by cotransfection with Gag-Pol and vesicular stomatitis virus G (VSV-G)-expressing vectors using the JetPEI transfection reagent (Polyplus). Viral supernatants were collected after 48 h and used for infection by spinoculation of Ramos and EBV-transformed lymphoblastoid B cells. After infection, green fluorescent protein (GFP)-positive cells were sorted (BD Influx, BD Bioscience) and grown for 1 week. Total RNA was extracted with TRIzol (Invitrogen) and converted into cDNA with SuperScript First-Strand Synthesis System (Invitrogen). Then, human *PAX5* expression was determined by quantitative real-time PCR (FastStart Universal SYBR Green Master Mix, Roche) using a 7500 Real-Time PCR system (Applied Biosystems). *GAPDH* was used as normalization control. The following primers were used: *PAX5* forward, 5′-GAGCGGGTGTGTGACAATGA-3′; *PAX5* reverse, 5′-GCACCGGAGACTCCTGAATAC-3′; *GAPDH* forward, 5′-GAAGGT GAAGGTCGGAGT-3′; and *GAPDH* reverse, 5′-GAAGATGGTGATGGGATTTC-3′.

To analyse the efficiency of the CRISPR/Cas9-induced deletions, DNA was extracted and *PAX5* enhancer was PCR-amplified using HotStarTaq DNA Polymerase (Qiagen) and *PAX5* enhancer-flanking oligonucleotides (forward) 5′-GTTGTCTTGGAGGACTTTCAG-3′, and (reverse) 5′- GTGTTATTGTGTATGTGGCAG-3′. To determine the presence of CRISPR/Cas9-induced mutations we performed heteroduplex cleavage assays using the Guide-it Mutation Detection Kit (Clontech) with primers (forward) 5′-AGGATGAGAACGGGCAAAC-3′ and (reverse) 5′-GGAGCTTCCAGCTGAACTGA-3′.

Statistical analysis. Fisher's test or non-parametric tests were used to correlate clinical and biological variables according to MBL or CLL, and the presence or absence of the different drivers herein analysed. We evaluated the clinical effect (TTT and overall survival) of all driver mutated genes and chromosomal regions with recurrent CNAs in 5 (1%) or more patients. TTT was evaluated only in patients with Binet A and B. TTT and overall survival curves from the date of sampling were plotted by the Kaplan–Meier method and compared by the log-rank test[57]. We examined separately the prognostic impact of point mutations in driver genes (substitutions or small indels) and CNAs. The clinical impact (TTT)

of *TP53*, *ATM* and *BIRC3* mutations was relatively similar to that of the loss of their respective chromosomal region, that is, del(17p) (*TP53*) and del(11q) (*ATM* and *BIRC3*), respectively (Extended Data Fig. 8). Therefore, to evaluate the prognostic impact for each gene/region, both types of alterations were combined. Although the clinical effect of deletions and mutations was somehow different for del(6q15)/*ZNF292* (Extended Data Fig. 8), owing to the fact that most point mutations in *ZNF292* were truncating, we also combined these two alterations to investigate the clinical effect. Finally, the number of cases with mutations or CNAs in the respective chromosomal region of 6p21/*NFKBIE*, 10q24/*NFKB2*, and 15q15/*MGA* was too small to perform a separate analysis and therefore we also combined both types of alterations. Multivariate Cox regression analysis was used to assess the independent prognostic impact from Binet stage and *IGHV* mutational status of each driver in the outcome of the patients. Proportional hazards were checked using Schoenfeld's test. We adjusted all the *P* values for multiple comparisons using the Benjamini–Hochberg correction. All statistical tests were two-sided and statistical significance was considered to be significant with an adjusted $P \leq 0.05$. All the analyses were performed using the SPSS 20 software (http://www.ibm.com) or R software v3.1.3.

Recurrently mutated genes in CLL were defined considering number and type of mutations, gene size and coverage, and local density of mutations derived from the 150 CLL/MBL WGS studies. To test whether a gene was mutated more frequently than expected by chance, we calculated the basal probability for each gene to suffer a non-synonymous mutation (P_{ns}) as: $P_{ns} = \dfrac{n_{ns}L\delta}{(n_{ns} + n_{s})E}$

In this equation, n_{ns} is the total number of possible non-synonymous mutations for this gene, n_{s} the total number of possible synonymous mutations, L is the effective length of the gene open reading frame (ORF), defined as the sum of the number of bases of the ORF for that gene which are callable at $10\times$ coverage for all exomes or whole genomes analysed, and E is the effective length of all coding regions analysed, defined as the sum of the total lengths of the coding regions that are callable at $10\times$ coverage for all exomes or whole genomes. Finally, δ is the local density of mutations for this locus, which is determined by dividing the number of somatic mutations identified in the 150 WGS studies analysed in a 0.5-Mb region centred on the gene of interest. Thus, the probability P to find M or more non-synonymous mutations in a given gene from a set of N total number of somatic mutations in all patients is:

$$P = 1 - \sum_{j=0}^{M-1} \binom{N}{j} p_{ns}^{j}(1 - p_{ns})^{N-j}$$

A score is computed by taking the base-10 logarithm of this probability (*P*). Genes for which more than 10% of somatic mutations caused a synonymous change were removed. Finally, 1,000 Monte–Carlo simulations were performed to estimate the FDR based on the total number of mutations observed (*N*), and the local mutational density for each gene. To identify genes that might be recurrently mutated in an IGHV subgroup, the same analysis was performed only with tumours belonging to the same group (IGHV-MUT or IGHV-UNMUT), and adjusting the local density of mutations for each subgroup according to the mutations obtained from WGS data. Genes were classified in three different tiers
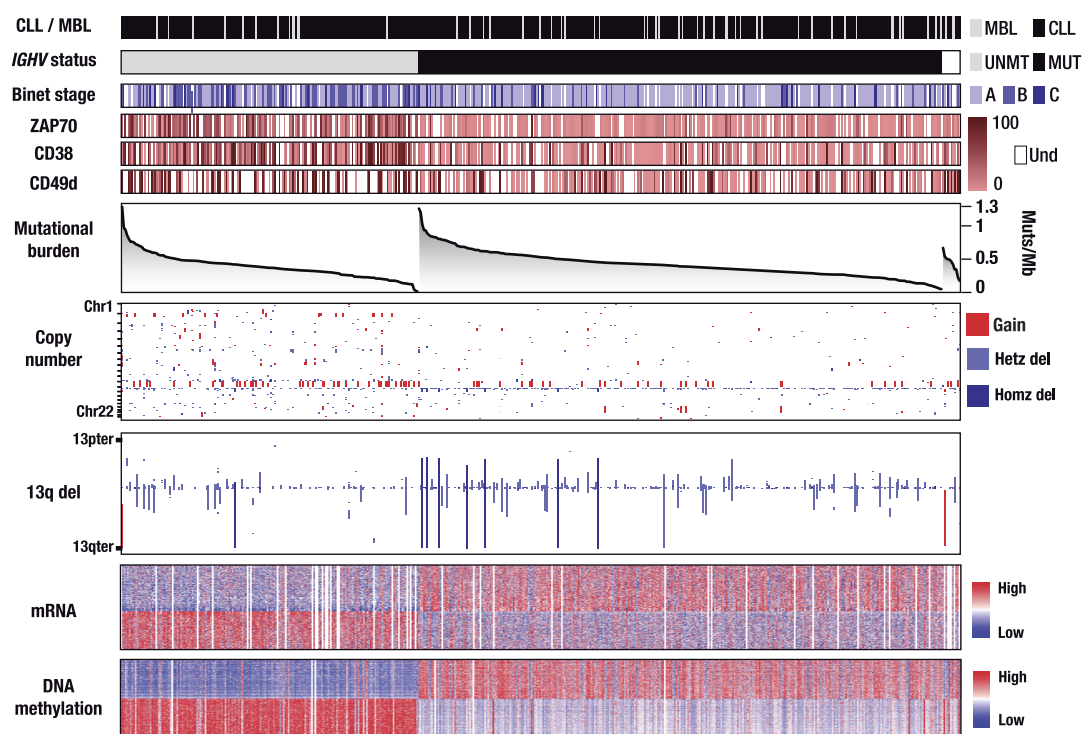
(Extended Data Table 2). Tier 1 corresponds to those genes that were identified as statistically mutated in CLL as described above. Tier 2 includes those genes that are not statistically mutated when analysing CLL, but appeared significant when only one subclass (IGHV-MUT or IGHV-UNMUT) was considered. In addition, genes showing either recurrent mutations affecting the same residue, or resulting in mainly loss-of-function mutations, were included in tier 2. Finally, genes classified in tier 3 include those genes that were not in tiers 1 or 2, but containing somatic mutations previously described as driver mutations in the literature.

A sample size of at least 500 tumours was selected during the ICGC study design, as this will give enough power to detect driver genes mutated in at least 3% of tumours[19].
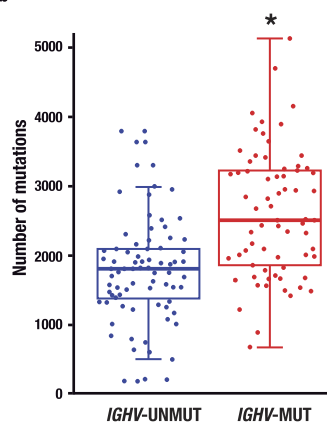
In silico prescription. Drugs with potential therapeutic interactions with driver oncogenic protein products were retrieved as described[37].

41. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
44. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475,** 101–105 (2011).
45. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3,** 246–259 (2013).
46. Delgado, J. *et al.* Genomic complexity and *IGHV* mutational status are key predictors of outcome of chronic lymphocytic leukemia patients with *TP53* disruption. *Haematologica* **99,** e231–e234 (2014).
47. Edelmann, J. *et al.* High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* **120,** 4783–4794 (2012).
48. Valdés-Mas, R., Bea, S., Puente, D. A., Lopez-Otin, C. & Puente, X. S. Estimation of copy number alterations from exome sequencing data. *PLoS ONE* **7,** e51422 (2012).
49. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98,** 288–295 (2011).
50. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium^R assay. *Epigenomics* **1,** 177–200 (2009).
51. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30,** 1363–1369 (2014).
52. Kluk, M. J. *et al.* Gauging NOTCH1 activation in cancer using immunohistochemistry. *PLoS ONE* **8,** e67306 (2013).
53. van de Werken, H. J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* **9,** 969–972 (2012).
54. van de Werken, H. J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513,** 89–112 (2012).
55. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnol.* **32,** 941–946 (2014).
56. Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nature Methods* **11,** 122–123 (2014).
57. Peto, R. & Pike, M. C. Conservatism of the approximation sigma (O-E)2-E in the logrank test for survival data or tumor incidence data. *Biometrics* **29,** 579–584 (1973).
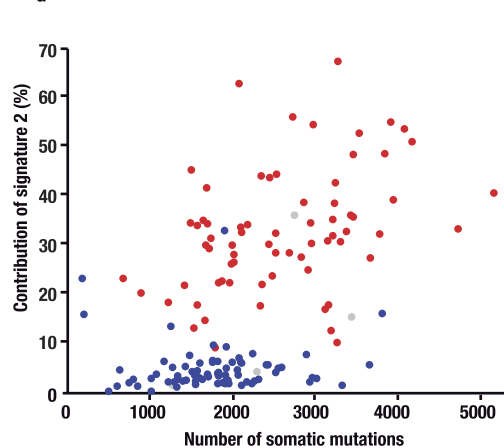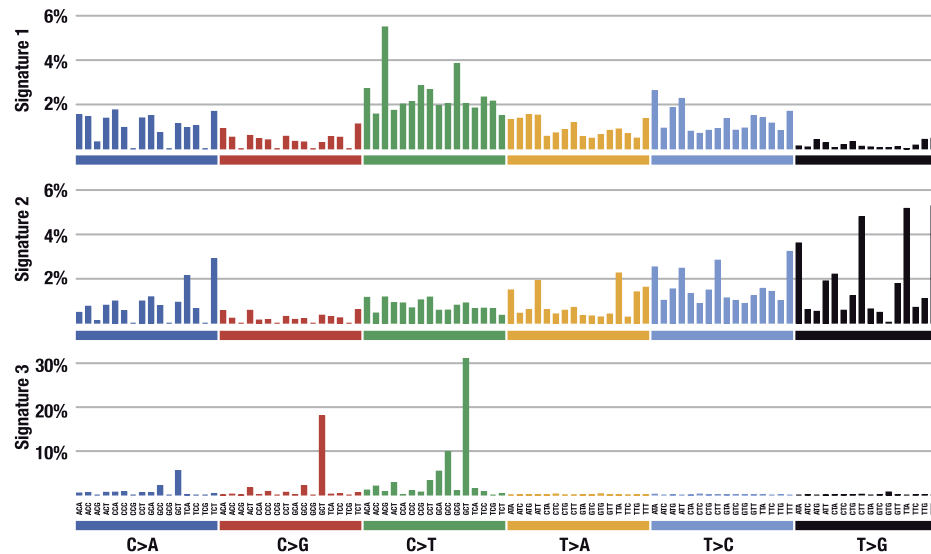
a

CLL / MBL — MBL / CLL

IGHV status — UNMT / MUT

Binet stage — A B C

ZAP70 — 100 / 0 / Und

CD38

CD49d

Mutational burden — 1.3 / 1 / 0.5 / 0 Muts/Mb

Copy number — Chr1 / Chr22 — Gain / Hetz del / Homz del

13q del — 13pter / 13qter

mRNA — High / Low

DNA methylation — High / Low

b

Number of mutations — *IGHV*-UNMUT / *IGHV*-MUT — *

d

Contribution of signature 2 (%) — Number of somatic mutations

c

Signature 1 — 6% / 4% / 2%

Signature 2 — 6% / 4% / 2%

Signature 3 — 30% / 20% / 10%

C>A   C>G   C>T   T>A   T>C   T>G

Extended Data Figure 1 | Molecular characterization of CLL and MBL subtypes. **a**, CLL and MBL cases are divided according to the somatic hypermutation mutational status of their clonotypic IGHV genes into IGHV-MUT (black) and IGHV-UNMUT (grey) subgroups. Clinical and molecular data from 506 cases profiled with four different platforms are shown. Chromosome 13 is shown in detail. Und, undetermined. **b**, Box plot showing the total number of somatic mutations identified in IGHV-MUT and IGHV-UNMUT cases by WGS (*$P < 3 \times 10^{-8}$). **c**, Main mutational signatures identified by WGS. **d**, Relationship between total number of mutations and contribution of signature 2 and the IGHV-status of tumours (red: IGHV-MUT; blue: IGHV-UNMUT; grey: undetermined).
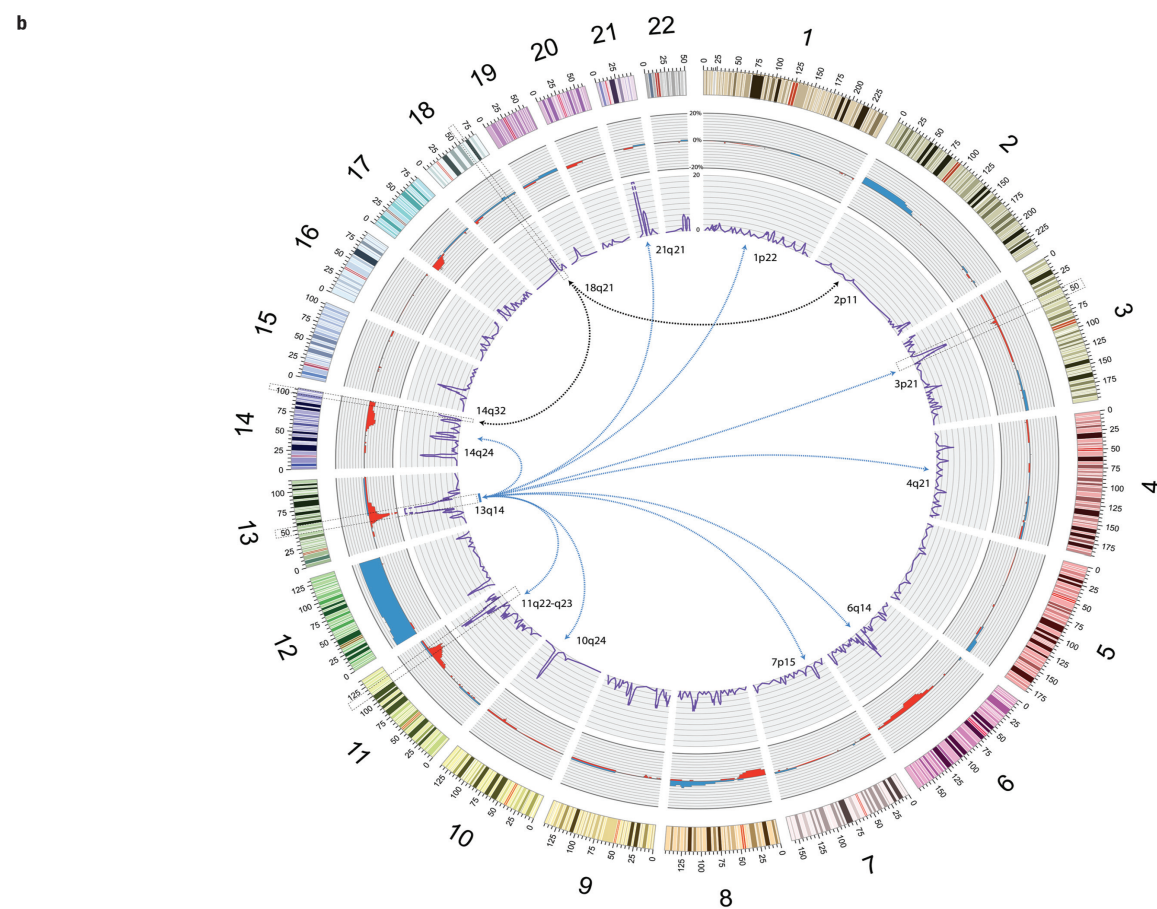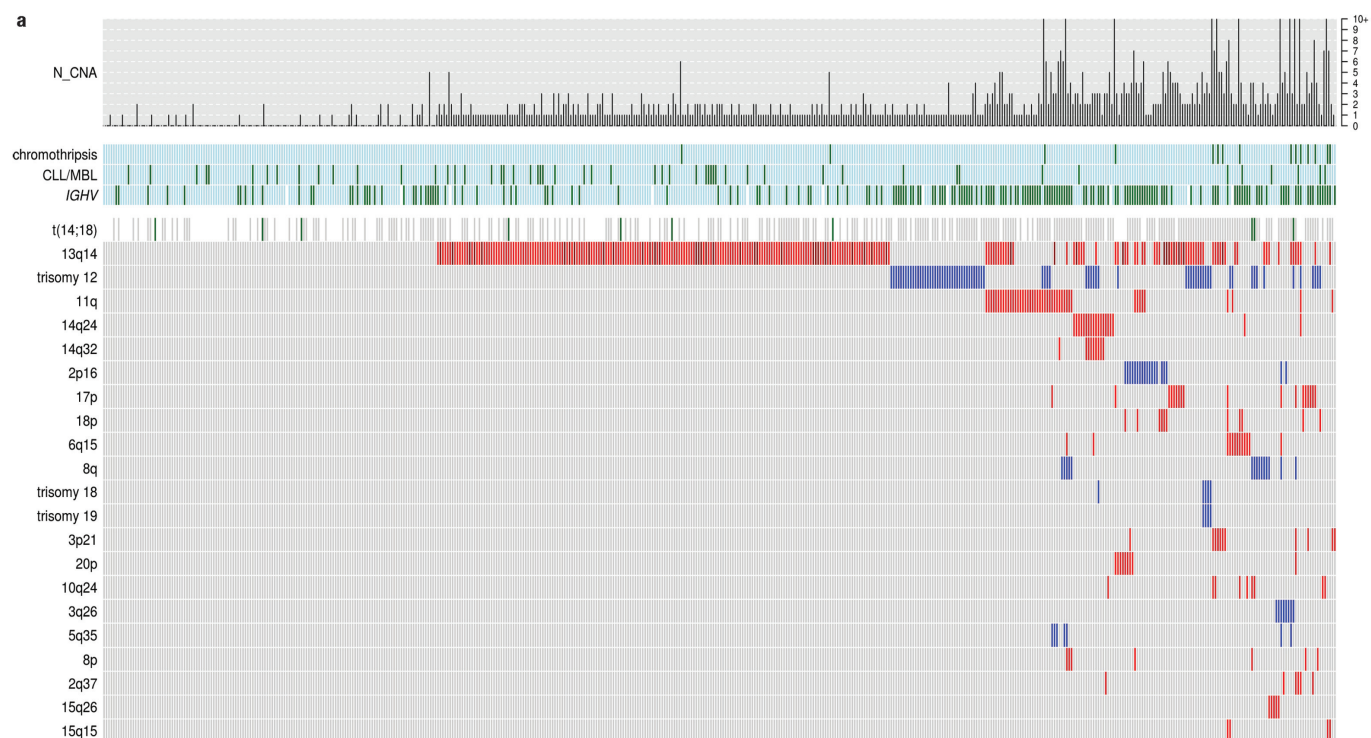
Extended Data Figure 2 | Distribution of CNAs and structural variants in 506 cases of CLL and MBL. a, The total number of CNAs detected per case is indicated on top. Clinicobiological characteristics of patients (CLL/MBL and IGHV status) are shown on the middle row (MBL and IGHV-UNMUT depicted with green lines), together with the presence of chromothripsis. The main DNA copy number alterations identified are shown on the bottom. The presence of a deletion is indicated by a red line, homozygous deletion by a Bordeaux colour, blue lines indicate the presence of a gain, translocation t(14;18) is shown in green, grey lines represent the absence of alteration, and white lines indicate that no information is available for the t(14;18) for that particular case. **b**, Circular diagram representation of the distribution of structural variants detected in 148 WGS CLL samples. Displayed in the outer layer we show recurrence in CNAs, followed by all the breakpoints derived from large (> 100 bp) intra- and inter-chromosomal rearrangements (dark blue) in the inner layer. For clarity, we have set the scale of CNAs to 20%, as the maximum, showing sequence gains and losses, as positive (blue) and negative (red) values, respectively. Rearrangements are displayed in absolute counts, indicating that the values in each of the regions do not reflect the recurrence among samples, as some regions with high values derive from one or two cases, normally with complex karyotypes. We highlighted with dashed squares those regions (3p21, 11q23, 13q14, 14q32 and 18q21) with rearrangements observed in more than 5% of cases with WGS. As to rearrangement events, of a total of 358 breakpoints were detected across all 148 samples, 41% of them correspond to interchromosomal translocations, while 59% occurred within chromosomes. Chromosomes 11 and 13 appear as the most rearranged, entailing 25% of all the breaks, followed by chromosomes 3 and 6 (with 8% each). Regarding interchromosomal rearrangements chromosomes 6, 8, 13 and 14 appear as the most translocated, being involved in 32% of all translocations observed. Recurrent breakpoints are indicated by arrows: black arrows for rearrangements affecting 18q21 and *BCL2* (four cases with 14q32 and one case with 2p11) and blue arrows for rearrangements affecting 13q14 (nine cases with different chromosomes).

a

IGK locus — Chr2

IGH locus — Chr14

Chr18

PROMOTER   5'UTR   EXON 1   EXON 2   3'UTR

BCL2

b

BCL2 expression (normalised array value)

1000

500

0

BCL2 wt (n=459)   BCL2 translocation (n=9)

c

IGH/BCL2

d

Patient	Chr	Pos	Chr	Pos
64	14	106357567	18	60793114
151	14	106329862	18	60788195
372	14	106382736	18	60793152
505	14	106330371	18	60786684
628	14	106346954	18	60788936
664	14	106135111	18	60884248
811	14	106357615	18	60788896
1271	2	89160678	18	60987856
1291	14	106350730	18	60793304
1423	14	106350741	18	60787269

e

CLL 16   CLL 141

CLL 753   CLL 294

CLL 880   CLL 853

f

der (X)

129.798 Mb (116%)

X : start - 147.426
2 : 208.928 - 243.199 (end)   GRM5

der (2)

238.688 Mb (99%)

2 : start - 207.230
11 : 567.964 - 567.975   DLG2
X : 148.328 - 148.335   CCDC81
8 : 111.408 - 86.863   APOA5
X : 148.375 - 155.270 (end)

der (8)

8 : start - 56.669
11 : 120.058 - 120.063   GRAMD1B-TENM4
11 : 119.621 - 119.647   OTOG
11 : 120.063 -123.462
11 : 78.963 - 63.962
11 : 17.645 - 15.782
11 : 119.413 - 119.170
11 : 119.062 -119.069
8 : 58.053 - 58.614   CADM1
11 : 115.218 - 115.698   CSMD3
8 : 113.636 - 113.021   DDX10
11 : 108.654 - 111.716   ALG9
8 : 111.522 - 111.795

129.798 Mb (89%)

8 : 68.748 - 86.732
8 : 116.716 - 146.364 (end)

der (11)

11 : start -12.316
11 : 27.578 - 38.693
11 : 88.438 - 92.907
11 : 56.024 - 50.015
11 : 119.170 - 119.164   CBL
11 : 133.001 - 132.122
11 : 22.791 - 19.100
11 : 85.165 - 86.129
11 : 116.664 - 119.062

11 : 93.270 - 107.891

79.331 Mb (60%)

X : 148.375 - 148.335
2 : 208.269 - 208.112
8 : 116.420 - 116.390
8 : 116.319 - 116.389
11 : 115.777 - 115.997   CUL5
11 : 107.894-107.891   NTM
11 : 60.999 - 56.802
2 : 208.9272 - 208.9278
11 : 132.105 - 125.100
11 : 114.433 - 114.151   PKNOX2
11 : 133.323 - 135.006 (en   OPCML

CLL 853

g

WCP8
WCP11

chr8
der(8)

der(11)

chr11

der(2)t(2;8)

CLL 853

Extended Data Figure 3 | Schematic view of the translocations involving *BCL2* and patterns of complex structural variants in the WGS of a CLL case. a, A total of nine translocations t(14;18)(q32;q21) were identified, resulting in the fusion of the *IGH* enhancer on the 3′ UTR of *BCL2*, as well as one translocation between the *IGK* locus on chromosome 2, t(2;18)(p11;q21), which affected the promoter region of *BCL2*. Cases with these translocations had multiple somatic mutations in the 5′-region of *BCL2* (arrowheads and lollipops). RNA-seq data analysis revealed an allelic imbalance, with the rearranged allele usually much more expressed that the germline allele (pie-charts within lollipops showing in red the mutant allele fraction detected by RNA-seq for each somatic mutation), probably reflecting the effect of the translocation on the expression of *BCL2* and recruitment of the SHM machinery to this locus. **b**, Gene expression analysis revealed that the presence of the t(14;18)(q32;q21) resulted in overexpression of *BCL2* in these cases when compared with other CLL or MBL cases. **c**, FISH analysis of CLL cells from case 151 using a dual colour fusion probe for *IGH* (green) and *BCL2* (red). Fusion signals are indicated with arrows. **d**, Case numbers and genomic coordinates for the detected translocations between immunoglobulin genes and *BCL2*. **e**, Circular representation of structural variants detected in six CLL tumours with complex rearrangements including four cases with chromoplexia (samples 16, 141, 294 and 753), chromothripsis (sample 880) and combined (sample 853). Chromosomes are represented in the outer layer, regions lost (red) and gained (blue) detected by SNP arrays are shown in the inner layer. Inter and intrachromosomal rearrangements are represented as black and blue lines, respectively. **f**, Reconstruction at base pair resolution of the resulting reorganized chromosomes in case 853 including der(X) in yellow, der(2) in dark blue, der(8) in green, and der(11) in red. In these reconstructions, only reorganized fragments larger than 100 bp are represented unless they involve interchromosomal translocations. Rearranged regions are not drawn to scale. Arrows denote inverted fragments relative to their normal and original orientation. Flanking portions of the derivative chromosomes without detected rearrangements are collapsed and shown as broken boxes. Estimated sizes (in Mb) for the resulting derivative chromosomes are shown on the left side, including the fraction (percentage) relative to the corresponding normal chromosome size. Asterisks indicate breakpoints that have been experimentally studied and verified. Genes disrupted by breakpoints are displayed on the left side of each of the proposed derivative chromosomes in purple. **g**, Whole-chromosome painting confirmed the sequencing reconstruction proposed in **b**. Simultaneous painting of chromosome 8 (green) and 11 (red) shows a normal chromosome 11 and a shorter chromosome der(11) as well as a normal chromosome 8 and der(8) that contains a fragment of chromosome 11 inserted below the centromeric region. In addition, a small fragment of chromosome 8 is detected in the telomeric region of derivative chromosome 2.

a

b

c
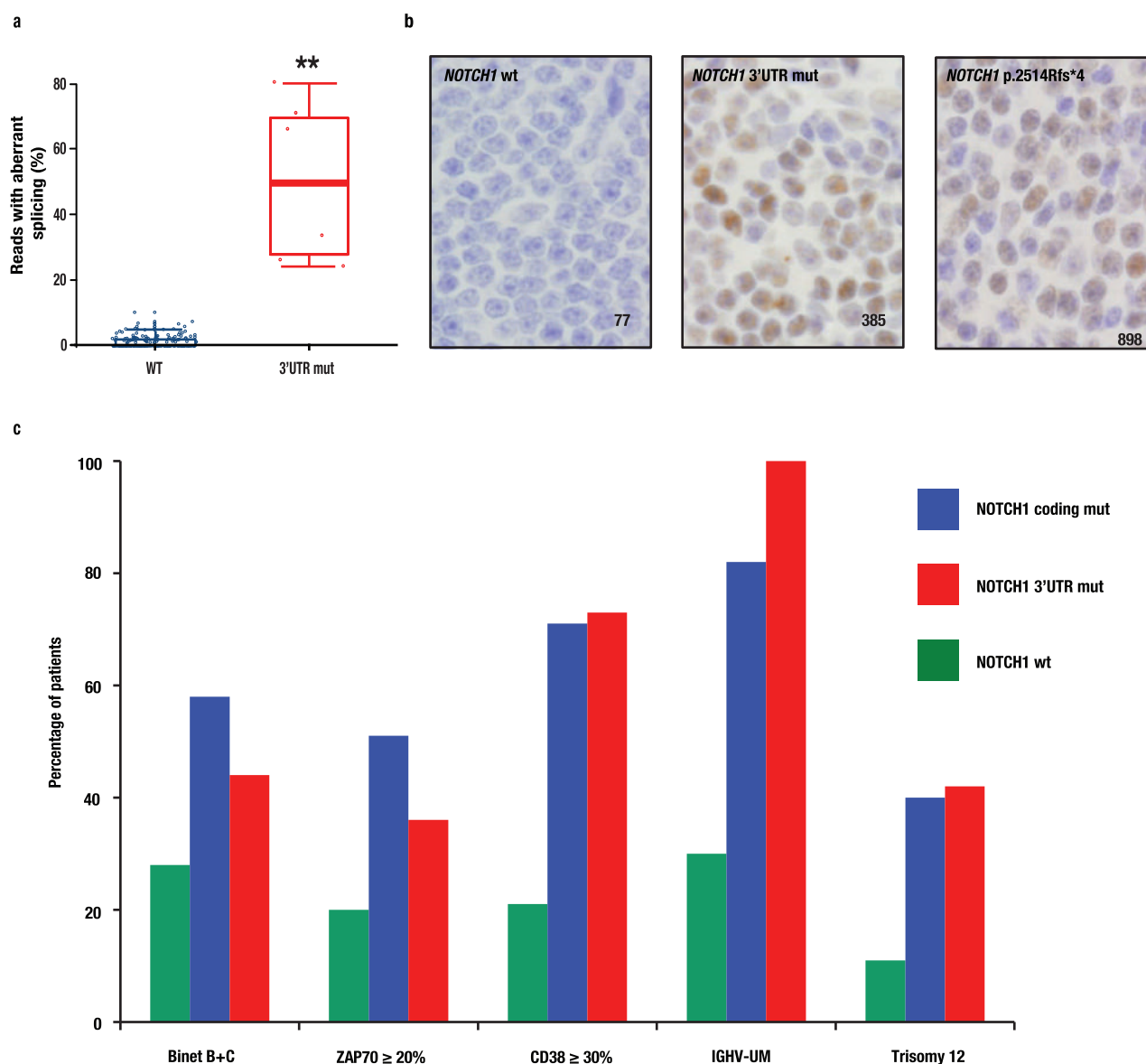
Extended Data Figure 4 | Effect of activating mutations in the 3′ UTR non-coding region of *NOTCH1*. **a**, Proportion of RNA-seq reads supporting an aberrant splicing in cases with or without mutations in the 3′ UTR of *NOTCH1* (**P** < 0.01). **b**, Immunohistochemistry of CLL cells with antibodies against NOTCH1 showing the nuclear accumulation of NOTCH1 in cells with coding or non-coding mutations in *NOTCH1* (case numbers are indicated inside). Original magnification, ×100. **c**, Clinical and biological features of patients with NOTCH1 mutations. Characteristics of patients with truncating mutations in the coding region of *NOTCH1*, in the 3′ UTR of *NOTCH1*, or without mutations in *NOTCH1* for Binet, ZAP70, CD38, IGHV status and trisomy 12.
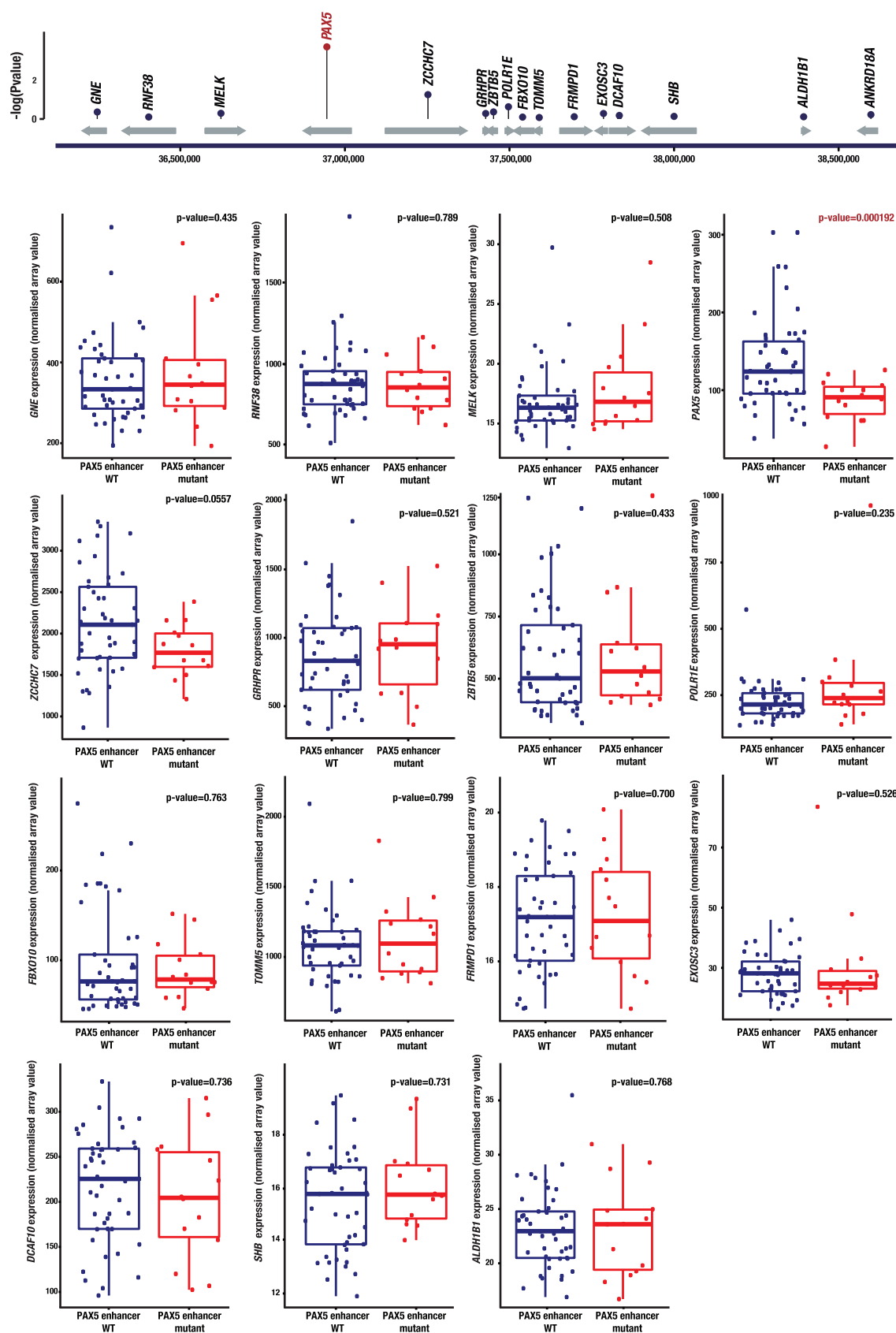
Extended Data Figure 5 | Effect of mutations in the PAX5 enhancer on gene expression. Comparative analysis of gene expression between IGHV-MUT CLL tumours with or without (WT) mutations in the PAX5 enhancer for 15 genes located around the recurrently mutated enhancer in CLL and MBL samples.
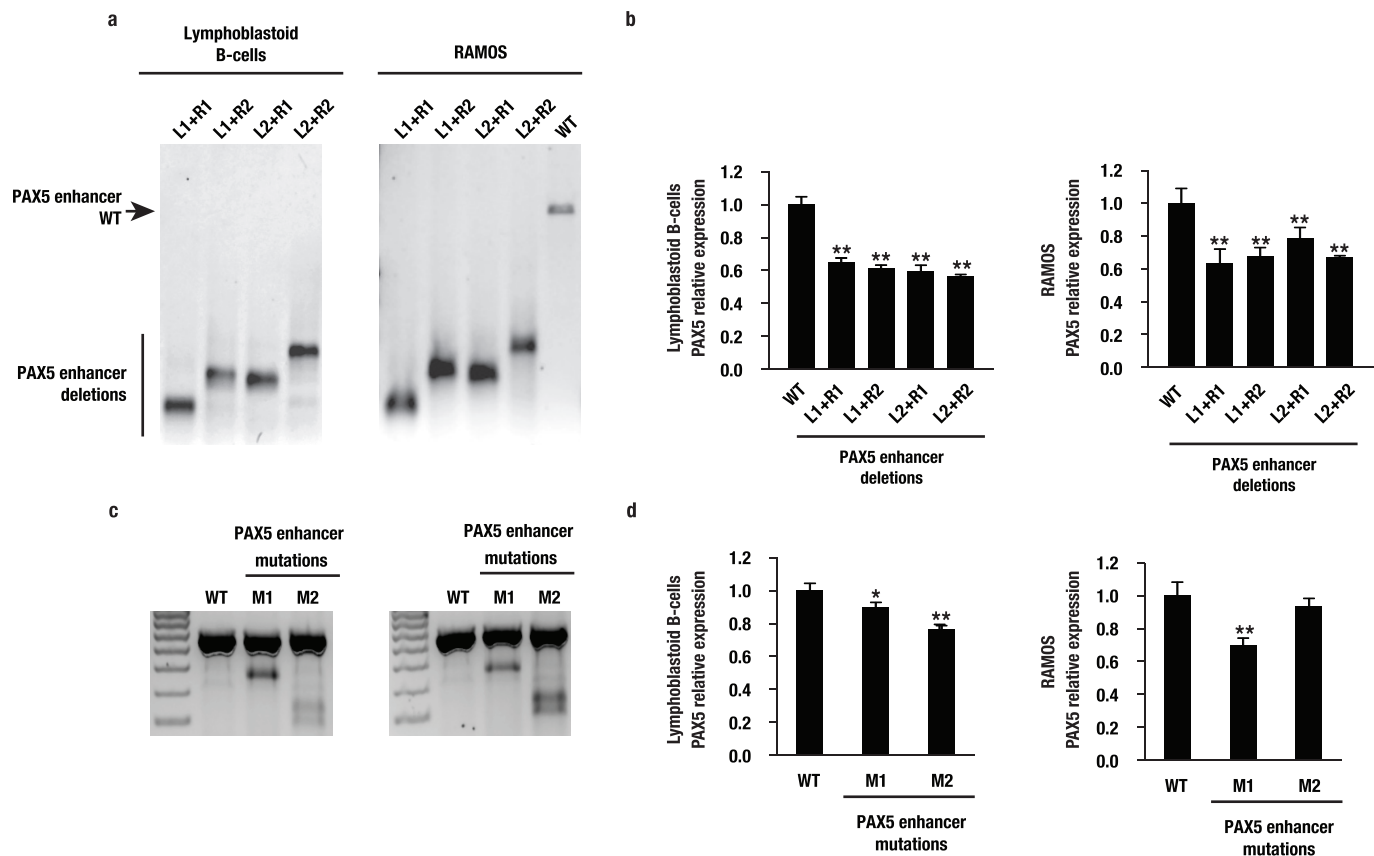
Extended Data Figure 6 | PAX5 enhancer deletion downregulates PAX5 expression in human B cell lines. a, PCR analysis of CRISPR/Cas9 deletion of PAX5 enhancer in lymphoblastoid B cells (left) and RAMOS cells (right). **b**, Quantitative RT–PCR (RT–qPCR) analysis of *PAX5* expression in PAX5 enhancer deleted lymphoblastoid B cells (left) and RAMOS cells (right). Bars represent mean relative *PAX5* mRNA levels after normalization to *GAPDH* expression and relative to wild-type cells. Errors bars represent the s.d. between technical triplicates of CRISPR/Cas9-induced mutations in PAX5 enhancer in lymphoblastoid B cells (left) and RAMOS cells (right). **c**, PCR analysis of CRISPR/Cas9-introduced mutations in the PAX5 enhancer in lymphoblastoid B cells (left) and RAMOS cells (right). **d**, RT–qPCR analysis of *PAX5* expression in PAX5-enhancer-mutated lymphoblastoid B cells (left) and RAMOS cells (right). Bars represent mean relative *PAX5* mRNA levels after normalization to *GAPDH* expression and relative to wild-type cells. Error bars represent the s.d. between technical triplicates (*$P < 0.05$; **$P < 0.01$).
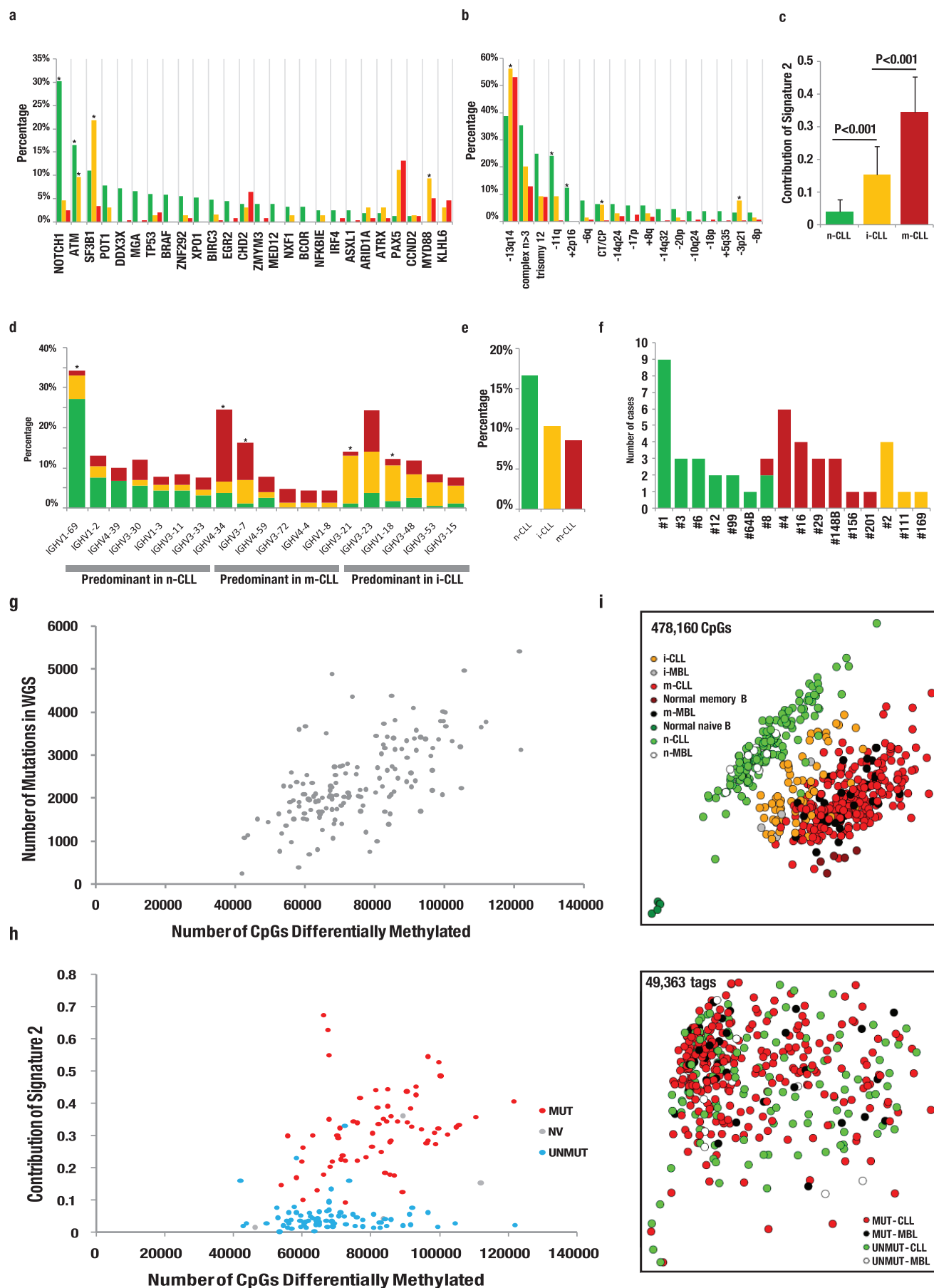
Extended Data Figure 7 | Distribution of genetic, epigenetic and expression features in CLL. Distribution of genetic features, family of IGHV rearrangements and BCR stereotypes in naive cell-like CLL cases, intermediate CLL and memory-cell-like CLL cases. **a**, Frequency of driver mutations. **b**, **c**, Copy number alterations (**b**) and contribution of signature 2 (**c**) according to the epigenetic classification (green: naive-like; red: memory-like; yellow: intermediate). MBL patients were excluded from this analysis. **d**, Usage of IGHV families. **e**, Proportion of cases with stereotyped IGHV sequences. **f**, Number of cases of each of the stereotyped subsets identified in our series. For the analysis shown in **e** and **f**, both CLL and MBL patients were merged. The asterisk on the top of the bars in **a** and **b** indicates that the frequency of the genetic feature is higher than expected by chance in one particular epigenetic subgroup ($P < 0.05$). CP, chromoplexy; CT, chromothripsis. **g**, Relationship between genetic and epigenetic alterations in CLL. Correlation between the total number of somatic mutations detected by WGS per case and the number of CpGs showing differential methylation per case as compared to naive B cells ($r = 0.64$, $P < 0.001$). **h**, Correlation between the contribution of signature 2 mutations and the number of differential CpGs as in **a**. Tumours are coloured according to their IGHV status. **i**, Comparative analysis of CLL and MBL. Principal component analysis of differential methylation (up) and gene expression (bottom) data derived from either CLL tumours or MBL samples, reveals that MBL samples usually clustered with their corresponding IGHV-status CLL samples.

TTT
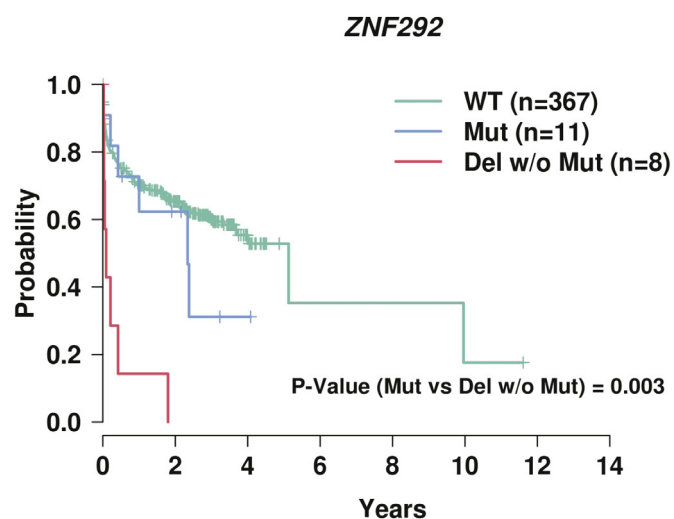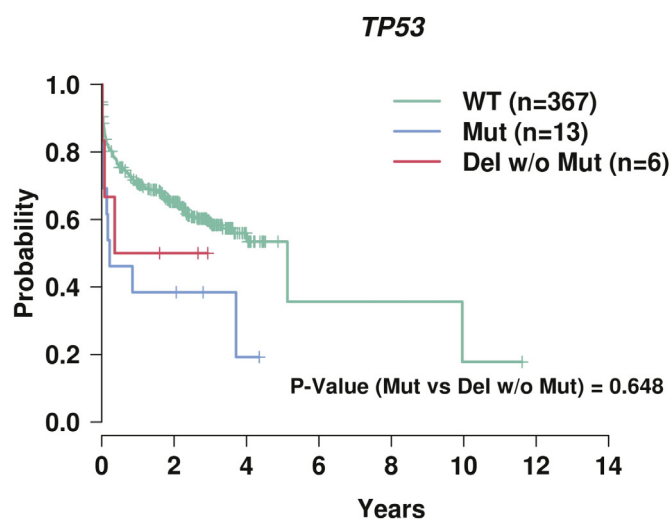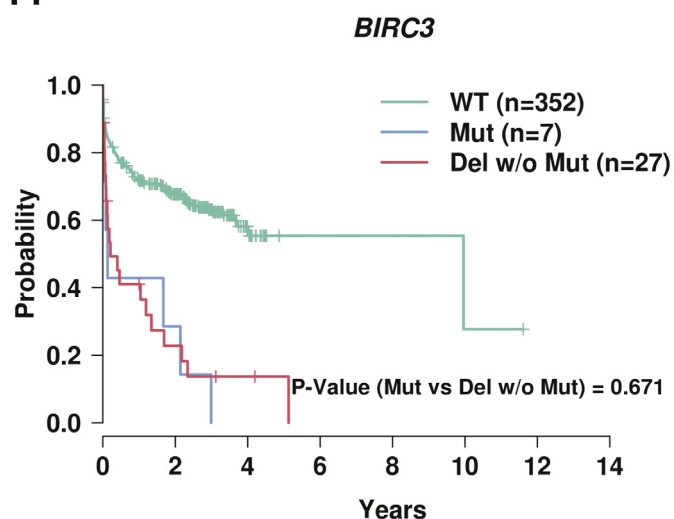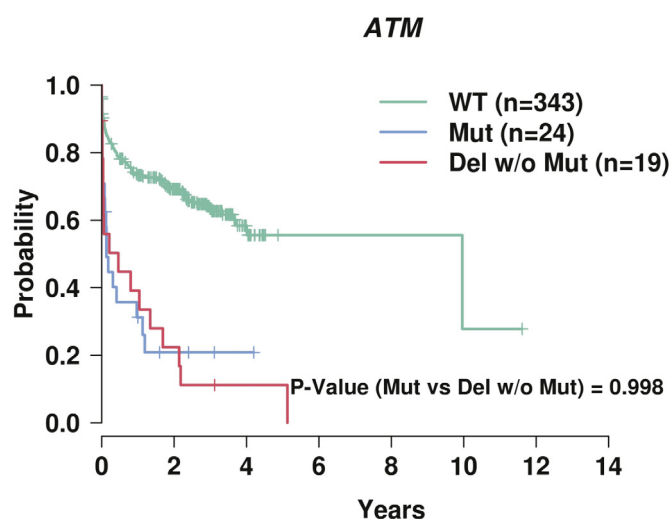
Extended Data Figure 8 | Kaplan–Meier plot of time to first treatment stratified by the type of aberration in *ATM, BIRC3, TP53* and *ZNF292* genes. TTT curves of the 386 untreated patients with Binet stage A or B. Cases are stratified according to the gene mutation status: wild type (green line), mutated and mutated+deleted (Mut, blue line) or deleted (Del, red line). The log-rank *P*-values comparing the mutated (blue line) and the deleted (red line) cases are shown.

Extended Data Table 1 | Clinical information at the time of sampling of 452 patients with CLL and 54 with MBL

Parameter	Category	CLL (n=452)	MBL[*] (n=54)
Gender	Male / Female	275 / 177	25 / 29
Age (years)		66 (19-93)	71 (39-89)
Clinical status	Diagnosis	92	13
	Stable disease	242	41
	Progression	118	---
Time from diagnosis to sampling (years)		3.12 (0-25.9)	3.68 (0-21.9)
Binet stage	A	297 (68%)	
	B	96 (22%)	---
	C	44 (10%)	
	Unknown	15	
Rai stage	0	206 (51%)	
	I-II	140 (35%)	---
	III-IV	58 (14%)	
	Unknown	48	
Lymphocytes (x10^9/L)		21.8 (1.69-300)	5.9 (3-11.2)
Absolute clonal B cells (x10^9/L)		14.3 (0.3†-209.7)	3.4 (1.1-4.99)
Hemoglobin (g/L)		135 (45-176)	142 (117-181)
Platelets (x10^9/L)		181 (7-791)	206 (57‡-394)
LDH	>UNL	48/376 (13%)	1/53 (2%)
CD38	High	114/430 (26%)	5/54 (9%)
ZAP-70	High	98/430 (23%)	14/54 (26%)
CD49d	High	100/423 (24%)	12/54 (22%)
Cytogenetics	Del13q	222 (49%)	31 (57%)
	Trisomy 12	68 (15%)	5 (9%)
	Del11q22	44 (10%)	2 (4%)
	Del17p	17 (4%)	1 (4%)
IGHV	Unmutated	166/445 (37%)	13/51 (25%)
Epigenetic subtype	Naïve-like	151/446 (34%)	11/54 (20%)
	Intermediate	64/446 (14%)	5/54 (9%)
	Memory-like	231/446 (52%)	38/54 (70%)
Follow-up, (years)	All surviving	2.9 (0.1-14.1)	2.1 (0.8-9.7)
5-year TTT (95% CI)	All	55% (46-64%)	2% (0-6%)
5-year OS (95% CI)	All	78% (71-85%)	96% (91-100%)

Sampling always before therapy; quantitative parameters are expressed as median (range); CD38 high: > 30% positive CLL cells; ZAP70 high: > 20% positive cells; IGHV-unmutated: > 98% identity with germ line. OS, overall survival.
*MBL only with CLL-like immunophenotype.
†Lower values corresponded to small lymphocytic lymphoma.
‡MBL with low platelet count due to chronic liver disease.

Extended Data Table 2 | Recurrently mutated genes in CLL and MBL by WGS, WES or CNAs

Symbol	Total CLL cases mutated	Score	Frequency CLL (n=452)	Frequency CLL IGHV-UNMUT	Frequency CLL IGHV-MUT	Frequency MBL (n=54)	Frequency MBL IGHV-UNMUT	Frequency MBL IGHV-MUT	Tier	Effect
NOTCH1	57	92.23	12.61	28.92	2.88	5.56	23.08	0.00	1	Activating
ATM	50	56.50	11.06	24.70	2.88	5.56	15.38	0.00	1	Truncating
SF3B1	39	56.47	8.63	13.25	6.12	3.70	7.69	2.63	1	Activating
BIRC3	40	56.41	8.85	19.28	2.52	3.70	7.69	2.63	1	Truncating
CHD2	27	81.96	5.97	3.61	7.55	7.41	7.69	7.89	1	Truncating
TP53	24	56.38	5.31	7.83	3.96	1.85	7.69	0.00	1	Truncating
ZNF292	21	56.95	4.65	10.24	1.08	7.41	30.77	0.00	1	Truncating
MYD88	18	66.46	3.98	0.00	6.47	0.00	0.00	0.00	1	Activating
KLHL6	14	63.49	3.10	0.00	5.04	0.00	0.00	0.00	1	-
POT1	14	44.11	3.10	8.43	0.00	7.41	23.07	2.63	1	Truncating
MGA	13	38.50	2.88	6.63	0.36	3.70	15.38	0.00	1	Truncating
DDX3X	12	38.99	2.65	5.42	0.72	0.00	0.00	0.00	1	Truncating
TRAF3	14	38.38	3.10	5.42	1.44	0.00	0.00	0.00	1	Truncating
SETD2	10	33.49	2.21	4.82	0.72	0.00	0.00	0.00	1	Truncating
BRAF	9	26.75	1.99	4.82	0.00	1.85	0.00	0.00	1	Activating
SYNE1	8	16.17	1.77	1.20	1.44	0.00	0.00	0.00	1	-
XPO1	8	24.49	1.77	4.82	0.00	1.85	7.69	0.00	1	Activating
IRF4	6	25.19	1.33	1.81	0.72	0.00	0.00	0.00	1	Activating
EGR2	9	27.83	1.99	4.22	0.72	0.00	0.00	0.00	1	Activating
CCND2	6	19.28	1.33	1.20	1.44	1.85	0.00	2.63	1	Activating
ZMYM3	8	19.43	1.77	3.61	0.72	0.00	0.00	0.00	1	Truncating
ARID1A	7	17.98	1.55	1.81	1.44	0.00	0.00	0.00	1	Truncating
ATRX	7	14.79	1.55	2.41	1.08	0.00	0.00	0.00	1	-
NFKBIE	5	18.95	1.11	2.41	0.36	0.00	7.69	0.00	1	Truncating
CNOT3	6	16.41	1.33	1.81	1.08	1.85	7.69	0.00	1	Activating
BCOR	6	14.99	1.33	3.61	0.00	0.00	0.00	0.00	1	Truncating
MED12	6	14.11	1.33	3.61	0.00	0.00	0.00	0.00	1	Activating
PTPN11	5	14.79	1.11	2.41	0.36	0.00	0.00	0.00	1	Activating
FBXW7	5	14.21	1.11	2.41	0.36	0.00	0.00	0.00	1	Activating
NXF1	6	13.99	1.33	2.41	0.72	0.00	0.00	0.00	1	Truncating
SETD1A	5	12.62	1.11	3.01	0.00	0.00	0.00	0.00	1	-
ASXL1	5	12.24	1.11	2.41	0.36	0.00	0.00	0.00	1	Truncating
FSIP2	7	13.86	1.55	1.20	1.44	0.00	0.00	0.00	1	Truncating
RPS15	4	14.62	0.88	2.41	0.00	0.00	0.00	0.00	1	Activating
FUBP1	4	14.33	0.88	2.41	0.00	0.00	0.00	0.00	1	Truncating
HIST1H1B	4	12.75	0.88	0.00	1.44	1.85	7.69	0.00	1	-
SPEN	6	11.48	1.33	1.81	1.08	0.00	0.00	0.00	2	Truncating
KIAA0947	5	10.97	1.11	0.60	1.44	0.00	0.00	0.00	2	Truncating
MLL2	5	10.77	1.11	0.00	1.80	0.00	0.00	0.00	2	Truncating
POLR3B	4	9.27	0.88	1.20	0.72	0.00	0.00	0.00	2	Activating
BAZ2A	3	6.57	0.66	1.20	0.36	0.00	0.00	0.00	2	Truncating
BAX	3	9.21	0.66	0.00	1.08	0.00	0.00	0.00	2	Truncating
KRAS	3	9.02	0.66	1.81	0.00	0.00	0.00	0.00	2	Activating
LUC7L2	3	8.71	0.66	0.60	0.72	0.00	0.00	0.00	2	Truncating
IKZF3	3	8.53	0.66	1.81	0.00	1.85	0.00	0.00	2	Activating
DNAJC11	3	8.22	0.66	1.20	0.36	0.00	0.00	0.00	2	Truncating
ZC3H18	3	7.38	0.66	0.60	0.72	0.00	0.00	0.00	2	Truncating
SKIV2L2	3	7.22	0.66	1.20	0.36	0.00	0.00	0.00	2	Activating
CREBBP	3	6.43	0.66	0.60	0.72	0.00	0.00	0.00	2	Truncating
ANKHD1	3	6.26	0.66	1.20	0.36	1.85	0.00	2.63	2	Truncating
NKAP	2	5.60	0.44	0.00	0.72	0.00	0.00	0.00	2	Truncating
TLR2	2	4.94	0.44	0.00	0.72	0.00	0.00	0.00	2	Activating
MED1	2	4.53	0.44	1.20	0.00	0.00	7.69	0.00	2	Activating
CDKN1B	1	-	0.22	0.00	0.36	0.00	0.00	0.00	3	Truncating
CDKN2A	1	-	0.22	0.00	0.36	0.00	0.00	0.00	3	Truncating
CD79A	1	-	0.22	0.00	0.36	0.00	0.00	0.00	3	Truncating
CD79B	1	-	0.22	0.00	0.36	0.00	0.00	0.00	3	Activating
IRAK1	1	-	0.22	0.00	0.36	0.00	0.00	0.00	3	Truncating
NRAS	1	-	0.22	0.60	0.00	0.00	0.00	0.00	3	Activating

ARTICLE

Mutations driving CLL and their evolution in progression and relapse

Dan A. Landau[1,2,3,4]*, Eugen Tausch[5]*, Amaro N. Taylor-Weiner[1]*, Chip Stewart[1], Johannes G. Reiter[1,2,6,7], Jasmin Bahlo[8], Sandra Kluth[8], Ivana Bozic[7,9], Mike Lawrence[1], Sebastian Böttcher[10], Scott L. Carter[1,11], Kristian Cibulskis[1], Daniel Mertens[5,12], Carrie L. Sougnez[1], Mara Rosenberg[1], Julian M. Hess[1], Jennifer Edelmann[5], Sabrina Kless[5], Michael Kneba[10], Matthias Ritgen[10], Anna Fink[8], Kirsten Fischer[8], Stacey Gabriel[1], Eric S. Lander[1], Martin A. Nowak[7,9,13], Hartmut Döhner[5], Michael Hallek[8,14]§, Donna Neuberg[15]§, Gad Getz[1,16]§, Stephan Stilgenbauer[5]§ & Catherine J. Wu[1,2,3,4]§

Which genetic alterations drive tumorigenesis and how they evolve over the course of disease and therapy are central questions in cancer biology. Here we identify 44 recurrently mutated genes and 11 recurrent somatic copy number variations through whole-exome sequencing of 538 chronic lymphocytic leukaemia (CLL) and matched germline DNA samples, 278 of which were collected in a prospective clinical trial. These include previously unrecognized putative cancer drivers (_RPS15_, _IKZF3_), and collectively identify RNA processing and export, MYC activity, and MAPK signalling as central pathways involved in CLL. Clonality analysis of this large data set further enabled reconstruction of temporal relationships between driver events. Direct comparison between matched pre-treatment and relapse samples from 59 patients demonstrated highly frequent clonal evolution. Thus, large sequencing data sets of clinically informative samples enable the discovery of novel genes associated with cancer, the network of relationships between the driver events, and their impact on disease relapse and clinical outcome.

In recent years, unbiased massively parallel sequencing of whole exomes (WES) in chronic lymphocytic leukaemia (CLL) has yielded fresh insights into the genetic basis of this disease[1–4]. Two important constraints have limited previous WES analyses. First, cohort size is critical for statistical inference of cancer drivers[5], and previous CLL WES series[3] had a power of only 68%, 23% and 7% to detect putative CLL genes mutated in 5%, 3% and 2% of patients, respectively (http://www.tumorportal.org/power)[5]. Limited cohort size has also curtailed the ability to effectively learn the relationships between CLL driver events, such as their co-occurrence and the temporal order of their acquisition. Second, the composition of the cohort of previous WES studies has limited the ability to accurately determine the impact of drivers and clonal heterogeneity on clinical outcome, since they included samples collected at variable times from subjects exposed to a variety of therapies.

To overcome these challenges, we analysed WES data from 538 CLLs, including 278 pre-treatment samples collected from subjects enrolled on the phase III CLL8 study[6]. This trial established the combination of fludarabine (F), cyclophosphamide (C) and rituximab (R) as the current standard-of-care first-line treatment for patients of good physical fitness, with a median of >6 years of follow-up. Here we report the discovery of novel genes associated with CLL, the comprehensive genetic characterization of samples from patients before exposure to a uniform and contemporary treatment, and the uncovering of features contributing to relapse from this therapy.

Unbiased candidate CLL gene discovery

We performed WES of CLL and matched germline samples, collected from 278 subjects enrolled on the CLL8 trial, with mean read depth of 95.0 and 95.7, respectively (Supplementary Tables 1 and 2). Consistent with previous CLL WES studies, we detected a mean ± s.d. rate of 21.5 ± 7.9 silent and non-silent single nucleotide variants (sSNVs) and somatic insertions and deletions (sIndels) per exome (Supplementary Tables 2 and 3)[1,3].

We inferred candidate cancer-associated genes in CLL through implementation of MutSig2CV[5,7]. To maximize statistical sensitivity for driver detection[5], we combined the CLL8 cohort with two previously reported and non-overlapping WES cohorts[1,3], thereby increasing the size of the cohort to 538 CLLs. This cohort size is expected to saturate candidate CLL gene discovery for genes mutated in 5% of patients, and provides 94% and 61% power to detect genes mutated in 3% and 2% of patients, respectively[5].

We detected 44 putative CLL driver genes, including 18 CLL mutated drivers that we previously identified[3], as well as 26 additional putative CLL genes (Figs 1 and 2 and Extended Data Figs 1 and 2). In total, 33.5% of CLLs harboured a mutation in at least one of these 26 additional genes. Targeted DNA sequencing as well as variant allele expression by RNA-seq demonstrated high rates of orthogonal validation (Extended Data Fig. 3).

Of the newly identified putative cancer-associated genes, some were previously suggested as CLL drivers in studies using other
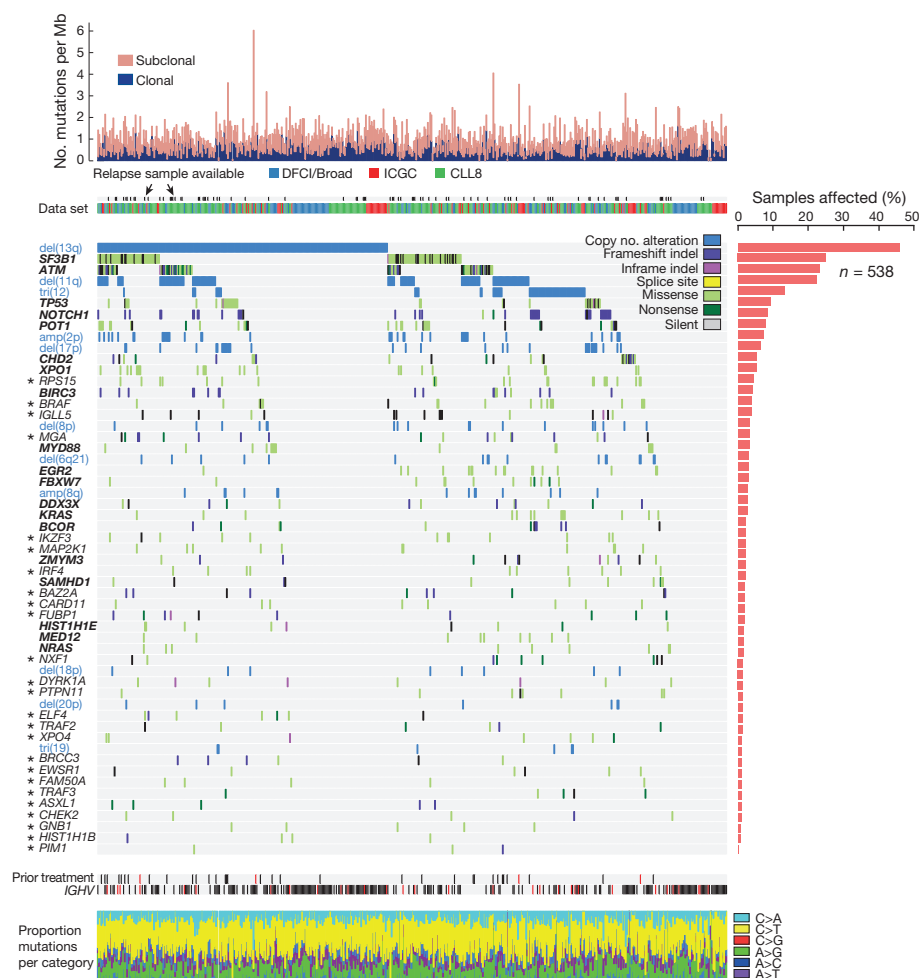
Figure 1 | The landscape of putative driver gene mutations and recurrent somatic copy number variations in CLL. Somatic mutation information is shown across the 55 putative driver genes and recurrent somatic copy number alterations (rows) for 538 primary patient samples (from CLL8 (green), Spanish ICGC (red) and DFCI/Broad (blue)) that underwent WES (columns). Blue labels, recurrent somatic CNAs; bold labels, putative CLL cancer genes previously identified in ref. 3; asterisked labels, additional cancer-associated genes identified in this study. Samples were annotated for *IGHV* status (black, mutated; white unmutated; red, unknown), and for exposure to therapy before sampling (black, previous therapy; white, no previous therapy; red, unknown previous treatment status).

detection platforms. For example, the suppressor of MYC *MGA* ($n = 17$, 3.2%), which we detected as recurrently inactivated by insertions and nonsense mutations, was previously found to be inactivated through deletions[8] and truncating mutations[8,9] in high-risk CLL

(Extended Data Fig. 4). A gene set enrichment analysis of matched RNA-seq data revealed downregulation of genes that are suppressed upon MYC activation in B cells[10] (Supplementary Table 4). In addition to *MGA*, we report two additional candidate driver genes that
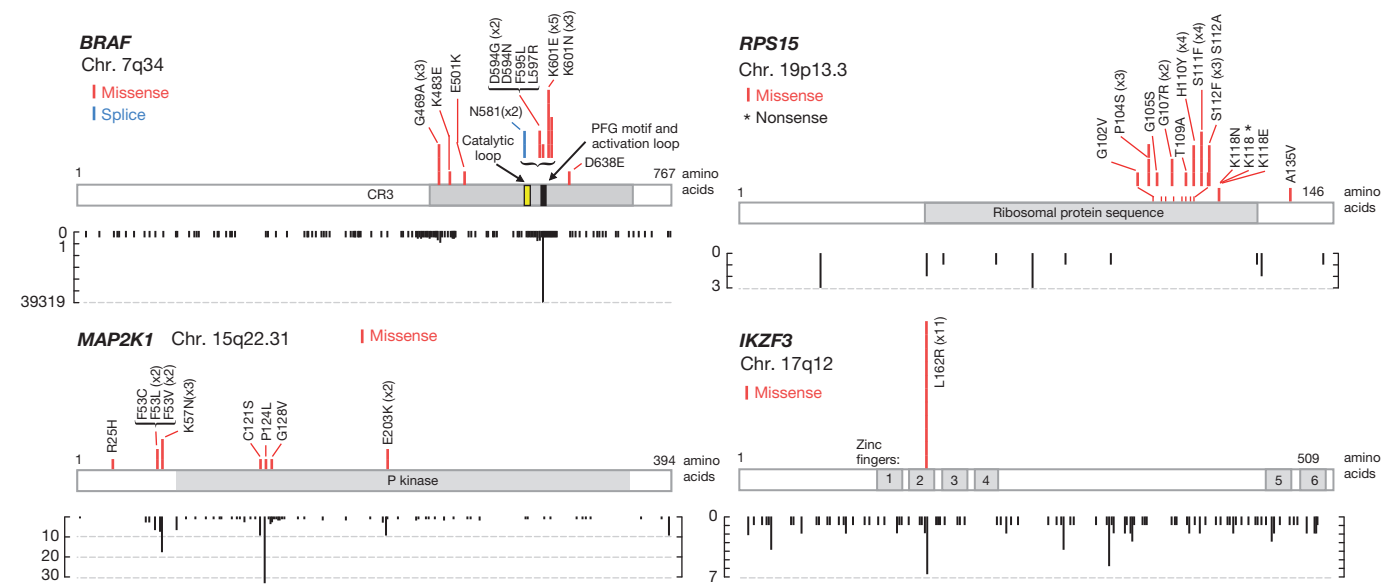
Figure 2 | Selected novel, putative driver gene maps. Individual gene mutation maps for select putative drivers, showing mutation subtype (for example, missense), position and evidence of mutational hotspots, based on

COSMIC database information (remaining gene maps shown in Extended Data Fig. 4). *y* axis counts at the bottom of the maps reflect the number of identified mutations in the COSMIC database.

probably modulate MYC activity (*PTPN11* (ref. 11) (*n* = 7, 1.3%) and *FUBP1* (ref. 12) (*n* = 9, 1.7%)), highlighting MYC-related proteins as drivers of CLL.

Another cellular process affected by novel CLL drivers is the MAPK–ERK pathway, with 8.7% of patients harbouring at least one mutation in CLL genes in this pathway. These included mutations in RAS genes (*NRAS*, *n* = 9 and *KRAS*, *n* = 14, totalling 4.1%); *BRAF* (*n* = 21, 3.7%); or the novel putative driver *MAP2K1* (*n* = 12, 2%). This finding suggests that further therapeutic exploration of MAPK–ERK pathway inhibitors in CLL would be beneficial. Notably, *BRAF* mutations in CLL did not involve the canonical hotspot (V600E) seen in other malignancies[5,13,14], but rather clustered heavily around the activation segment of the kinase domain (Fig. 2). This may be indicative of a different mechanism of activity[15,16], and has clinical implications, as BRAF inhibitors are thought to be less effective for non-canonical BRAF mutations[17,18].

In addition to highlighting novel cellular processes and pathways affected in CLL, many of the 26 additional CLL genes more densely annotated pathways or functional categories previously identified in CLL[19], including RNA processing and export (*FUBP1*, *XPO4*, *EWSR1* and *NXF1*), DNA damage (*CHEK2*, *BRCC3*, *ELF4* (ref. 20) and *DYRK1A* (ref. 21)), chromatin modification (*ASXL1*, *HIST1H1B*, *BAZ2B* and *IKZF3*) and B-cell-activity-related pathways (*TRAF2*, *TRAF3* and *CARD11*).

We discovered a number of putative CLL drivers previously unrecognized in human cancer. In a first example, we found that *RPS15* was recurrently mutated (*n* = 23, 4.3%), with mutations localized to the carboxy-terminal region (Fig. 2) at highly conserved sites (median conservation score of 94 out of 100). This component of the S40 ribosomal subunit has not been extensively studied in cancer, although rare mutations have been identified in Diamond–Blackfan anaemia[22]. A gene set enrichment analysis revealed upregulation of gene sets related to adverse outcome in CLL as well as immune response gene sets (Supplementary Table 4). In another example of a previously unrecognized cancer gene, we identified recurrent L162R substitutions (*n* = 11, 2.0%) in *IKZF3*, targeting a highly conserved amino acid (93 out of 100 conservation score). This gene is a key transcription factor in B-cell development[23], and its upregulation has been associated with adverse outcome[24,25].

In addition to sSNVs and sIndels, we characterized somatic copy number alterations (CNAs) directly from the WES data (Extended Data Fig. 5 and Supplementary Tables 5 and 6). When we accounted for all 55 identified driver events—including non-silent sSNVs and sIndels in putative CLL genes (*n* = 44), and recurrent somatic CNAs (*n* = 11)—91.1% of CLLs contained at least one driver. Moreover, 65.4% of CLLs now harboured at least 2 drivers, and 44.4% at least 3 drivers, compared with 55.9% and 31.8% were we to exclude the 26 additional CLL genes.

Drivers and CLL characteristics

The larger cohort size also provided statistical power to examine associations between genetic alterations and key CLL features. First, we examined whether mutations differed between *IGHV* mutated and unmutated subtypes, the two main subtypes of CLL. In agreement with the relative clinical aggressiveness of *IGHV* unmutated CLL, most drivers were found in a higher proportion in this subtype (Extended Data Fig. 6a). Only three driver genes were enriched in the *IGHV* mutated CLL (del(13q), *MYD88* and *CHD2*), suggesting a role for these specific alterations within the oncogenic process of this subtype.

Second, since therapy could lead to selection of particular driver events, we examined the 33 samples (6.2%, none enrolled on CLL8) that had received therapy before sampling. Previous treatment was associated with enrichment in *TP53* and *BIRC3* mutations del(17p) and del(11q), as previously indicated[26], as well as in mutated *DDX3X* and *MAP2K1*, suggesting their selection by therapeutic interventions (Extended Data Fig. 6b).

Third, we examined whether coherent patterns of co-occurrence of driver events were evident, limiting our analysis to the 31 drivers with >10 affected patients. Of 465 possible pairs, 11 combinations had statistically significant high or low co-occurrence (Extended Data Fig. 6c, d). As expected, a high degree of co-occurrence was found between mutated *TP53* and del(17p), and between mutated *ATM* and del(11q). Both mutated *ATM* and del(11q) significantly co-occurred with amp(2p), and associations between the presence of tri(12) with mutated *BIRC3* and with mutated *BCOR* were also found. A significantly low rate of co-occurrence was seen between del(13q) and tri(12).

Fourth, we examined the temporal sequence of driver acquisition in the evolutionary history of CLL. To do this, we computed the cancer-cell fraction (CCF) of each mutation across the 538 samples, and identified mutations as either clonal or subclonal[27] (58.1% of mutations classified as subclonal). Both clonal and subclonal sSNVs were similarly dominated by C > T transitions at CpG sites (Extended Data Fig. 7).

We first classified driver events probably acquired earlier or later in the disease course based on the proportion of cases in which the driver was found as clonal (Fig. 3a). This large data set further enabled the inference of temporal relationships between pairs of drivers. We systematically identified instances in which a clonal driver was found together with a subclonal driver within the same sample, as these pairs reflect the acquisition of one lesion (clonal) followed by another (subclonal), providing a temporal 'edge' leading from the former to the latter[28,29]. For each driver, we calculated the relative enrichment of out-going edges compared to in-going edges to define early, late and intermediary drivers (Supplementary Table 7). For 23 pairs connected by at least 5 edges, we further established the temporal relationship between the two drivers in each pair, and thereby constructed a temporal map of the evolutionary trajectories of CLL (Supplementary Table 8 and Fig. 3b). This network highlights somatic CNAs as the earliest events with two distinct points of departure involving del(13q) and tri(12). It further demonstrates an early convergence towards del(11q) and substantial diversity in late drivers. Finally, this analysis suggests that in the case of the tumour suppressor genes *ATM* and *BIRC3*, copy loss precedes sSNVs and sIndels in biallelic inactivation.

Impact on clinical outcome

We examined whether the presence of any of the drivers detected in at least 10 of the 278 pre-treatment CLL8 samples was associated with impact on clinical outcome (Fig. 4a and Extended Data Figs 8 and 9; the genomics analysis team was blinded to the clinical outcome data). Previous investigations suggested an impact for 7 CLL genes (*SF3B1*, *ATM*, *TP53*, *XPO1*, *EGR2*, *POT1* and *BIRC3*)[30–33]. We found shorter progression-free survival (PFS) associated only with *TP53* and *SF3B1* mutations. Of the newly identified recurrent lesions evaluated (*MGA*, *BRAF* and *RPS15*), we observed a shorter PFS with mutated *RPS15* (Bonferroni *P* = 0.024).

The presence of a detectable pre-treatment subclonal driver has been previously associated with shorter remissions in patients treated with heterogeneous therapies[3]. In the CLL8 cohort, we again found that the presence of a pre-treatment subclonal driver was associated with a significantly shorter PFS (hazard ratio (HR) 1.6 (95% confidence interval (CI) 1.2–2.2), *P* = 0.004). This association remained significant in both the FC (fludarabine and cyclophosphamide) and FCR (fludarabine, cyclophosphamide and rituximab) treatment arms (Fig. 4b), with a non-significant trend when *IGHV* mutation status was added to a multivariable model in addition to the treatment arm (1.3 (0.9–1.9), *P* = 0.102).

Clonal evolution at disease relapse

To define clonal evolution in disease relapse, we performed WES on matched samples collected at the time of relapse from 59 of 278 CLL8
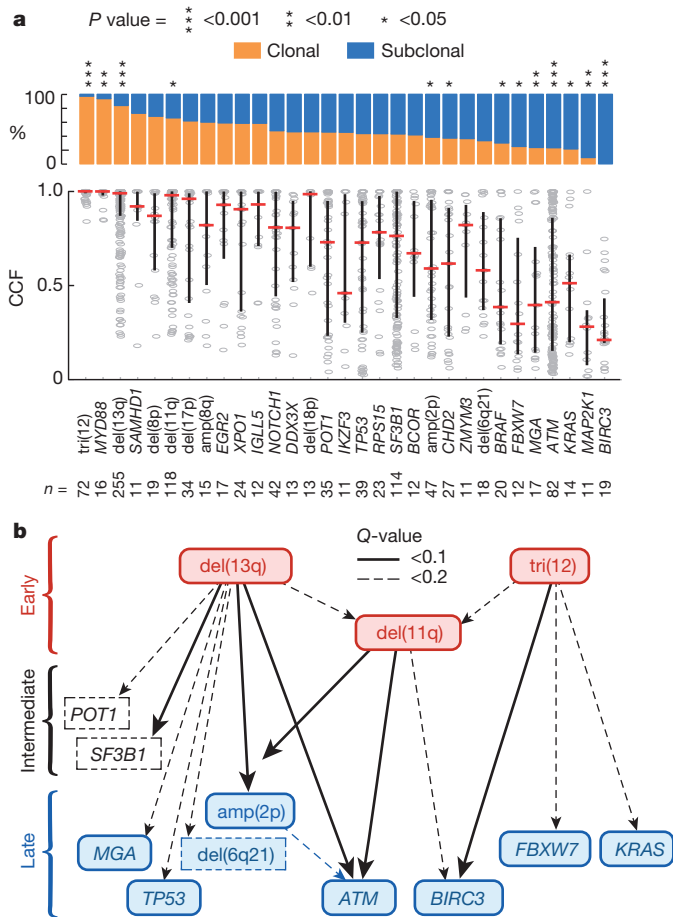
Figure 3 | Inferred evolutionary history of CLL. a, The proportion in which a recurrent driver is found as clonal or subclonal across the 538 samples is provided (top), along with the individual cancer cell fraction (CCF) values for each sample affected by a driver (tested for each driver with a Fisher's exact test, comparing to the cumulative proportions of clonal and subclonal drivers excluding the driver evaluated). Median CCF values are shown (bottom, bars represent the median and interquartile range for each driver). **b,** Temporally direct edges are drawn when two drivers are found in the same sample, one in clonal and the other in subclonal frequency. These edges are used to infer the temporal sequences in CLL evolution, leading from early, through intermediate to late drivers. Note that only driver pairs with at least five connecting edges were tested for statistical significance and only drivers connected by at least one statistically significant edge are displayed (see Supplementary Methods and Supplementary Tables 6 and 7).

Figure 4 | Associations of CLL drivers with clinical outcome. a, Kaplan–Meier analysis (with logrank *P* values) for putative drivers with associated impact on progression-free survival (PFS) or overall survival (OS) probabilities in the cohort of 278 patients that were treated as part of the CLL8 trial. For candidate CLL genes tested here for the first time regarding impact on outcome, a Bonferroni *P* value is also shown. **b,** Presence of a subclonal driver is associated with a lower probability of PFS, in both the FC and FCR arms, and a trend towards shorter OS.

subjects (Supplementary Tables 9 and 10). We observed large clonal shifts between pre-treatment and relapse samples in the majority of cases (57 of 59), thus demonstrating that CLL evolution after therapy is the rule rather than the exception (Fig. 5a). The relapse clone was already detectable in pre-treatment WES in 18 of 59 (30%) cases, demonstrating that the study of pre-treatment diversity anticipates the future evolutionary trajectories of the relapsed disease[34]. By targeted deep sequencing, we screened for relapse drivers in 11 of the 41 of pre-treatment samples in which WES did not detect the relapse driver. In 7 of these 11 CLLs, at least one relapse driver was detected in the pre-treatment sample (Supplementary Table 10).

We further compared the pre-treatment and relapse CCF for each driver, and observed three general patterns. First, tri(12), del(13q) and del(11q), suggested as early drivers (Fig. 3b), tended to remain stably clonal despite marked, often branched, evolution (Fig. 5b (CLL cases GCLL-115 and GCLL-307), Fig. 5c, top row, and Extended Data Fig. 10). This confirms that these are indeed early events probably shared by the entire malignant population. Second, *TP53* mutations and del(17p) demonstrated increases in CCF upon relapse, suggesting

a fitness advantage under therapeutic selection (Fig. 5b (GCLL-27) and Fig. 5c, middle row). The novel driver *IKZF3* increased in CCF in 3 of 4 relapse cases (and remained clonal in the fourth), supporting the suggestion that these mutations probably enhance fitness. Third, mutations in *SF3B1* and *ATM*, identified as temporally intermediate or late drivers, seemed just as likely to decline in CCF as they were to increase (Fig. 5c, bottom row). These results suggest that within this therapeutic context such mutations do not provide the same strength of fitness advantage compared to *TP53* disruption. In addition, we observed nine instances each of multiple distinct alleles of *ATM* and *SF3B1* mutations within the same CLL (for example, GCLL-307 in Fig. 5b), indicating convergent evolution of these late-occurring CLL drivers.

This series also informs us regarding the mutagenesis of the tumour suppressor genes *TP53* and *ATM*, where biallelic inactivation is common. In the case of *ATM*, we typically find a fixed clonal del(11q22.3) and subclones harbouring sSNVs affecting the other allele that shift in CCF over time (for example, GCLL-307). We confirmed that the breakpoints of somatic CNAs in matched relapse and pre-treatment samples were highly consistent, probably representing the same deletion event. These data suggest that mono-allelic *ATM* deletion provides a fitness advantage that enables the expansion of the malignant population with subsequent growth of multiple co-existing clones
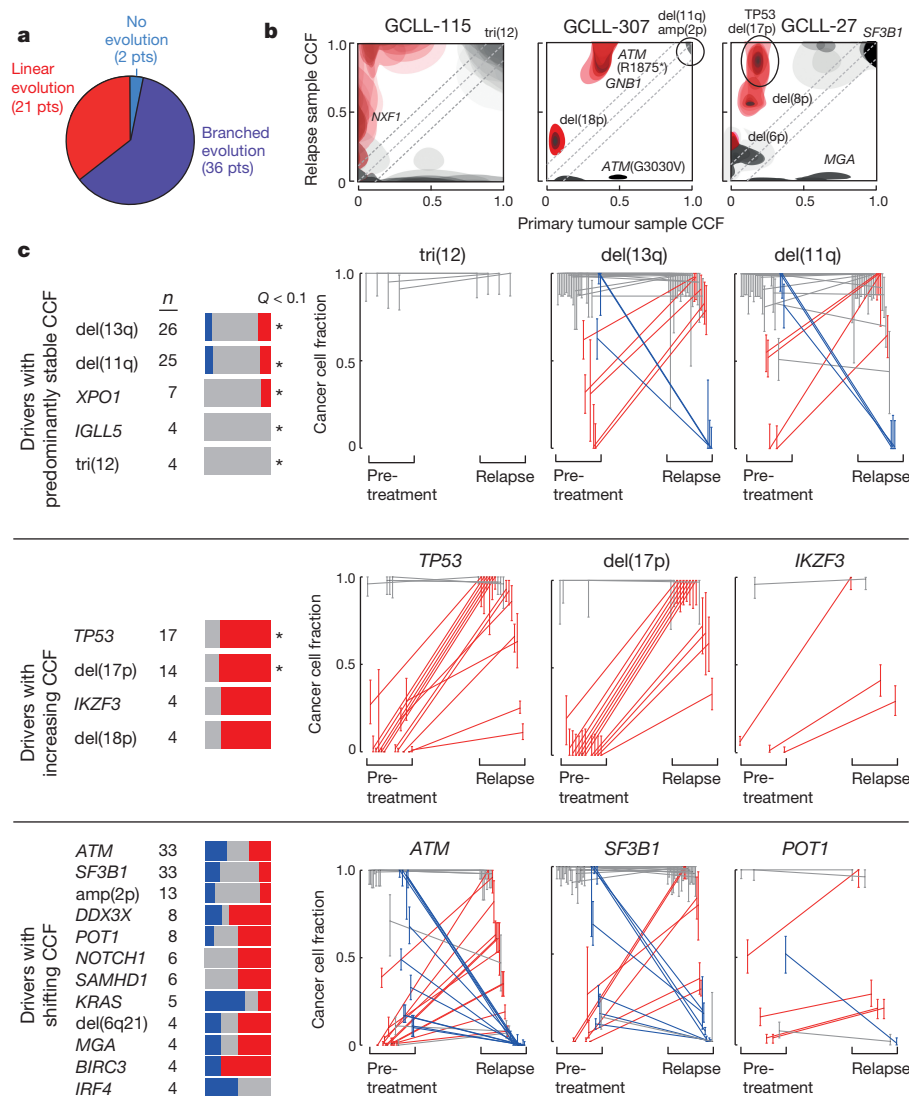
Figure 5 | Matched pre-treatment and relapse samples reveal patterns of clonal evolution in relation to therapy. a, The number and proportion of the patterns of clonal evolution of CLLs studied at pre-treatment and at relapse. pts, patients. **b**, Selected plots of 2D clustering of pre-treatment and relapse CCF demonstrating clonal stability of tri(12) (CLL case GCLL-115), concordant increase in CCFs of *TP53* and del(17p) (GCLL-27), and clonal shifts in *ATM* sSNVs in a sample with clonally stable monoallelic deletion of *ATM* (GCLL-307). Red colouring was added when greater than half of the CCF probability indicated>0.1 increase in CCF at relapse. **c**, Clonal evolution of CLL drivers. Left panel: for each driver with at least 4 instances detected across the 59 CLLs, the proportion of instances where the CCF increased (red), decreased (blue) or remained stable (grey) over time is shown (see Supplementary Methods for details of the statistical analysis). The driver events were distributed to three main groups: predominately stable events (top); predominately increasing CCF (middle); and all other patterns (bottom). Right panel: comparison (modal CCF with 95% CI) between pre-treatment and relapse samples for select CLL drivers (see Extended Data Fig. 10 for the remaining driver events from the cohort of 59 CLLs).

that harbour a 'second hit' (genetic disruption of the remaining allele). Thus, while a biallelic lesion is clearly selected for (Extended Data Fig. 6c), the longitudinal data support the temporal analysis (Fig. 3b) in which del(11q) precedes *ATM* mutations, reflecting the higher likelihood of a focal copy number loss compared with a deleterious point mutation[35,36]. In contrast, we consistently observed a concordant rise of del(17p) and *TP53* mutations in all 12 CLLs harbouring both of these events, and none of these cases exhibited multiple distinctly evolving *TP53* mutated clones. These observations suggest that a true biallelic inactivation of *TP53* is required, and indeed, across the 538 CLL samples, the odds ratio for co-occurrence of del(17p) and *TP53* mutation was far greater than the odds ratio for co-occurrence of del(11q) and *ATM* mutation (97.22 versus 10.99, respectively). These observations are in agreement with a recent analysis that suggested that with the exception of a few genes such as *TP53*, tumour suppressor genes in sporadic cancers are haploinsufficient to begin with, and that the second hit only further builds on this fitness advantage[37].

Conclusions

This study of WES in CLL enabled a comprehensive identification of putative cancer-associated genes in CLL, generating novel hypotheses regarding the biology of this disease, and identifying previously unrecognized putative CLL drivers such as *RPS15* and *IKZF3*. The detailed characterization of the compendium of driver lesions in cancer is of

particular importance as we strive to develop personalized medicine, because driver genes may inform prognosis (for example, *RPS15* mutations) and identify lesions that may be targeted by therapeutic intervention (for example, MAPK pathway mutations and specifically the unexpected enrichment for non-canonical *BRAF* mutations). Through the inclusion of samples collected within a landmark clinical trial with mature outcome data, we could further study the impact of genetic alterations in the context of the current standard-of-care front-line therapy. As targeted therapy is rapidly transforming the treatment algorithms for CLL, future studies will be required to re-examine these associations in this context[38].

An important benefit of the larger cohort size is the enhanced ability to explore relationships between driver lesions based on patterns of their co-occurrence. Focusing on temporal patterns of driver acquisition—based on the distinction between clonal versus subclonal alterations in a cross-sectional analysis—we derived a temporal map for the evolutionary history of CLL. In the context of relapse after first-line fludarabine-based therapy, we note highly frequent clonal evolution, and that the future evolutionary trajectories were already anticipated in the pre-treatment sample in one-third of cases with WES.

This study provides an indication of the potential benefits to be gained by applying novel genomic technologies to growing cohort sizes across leukaemias: the continued discovery of novel candidate cancer genes, the deeper integration of genetic analysis with standardized clinical information (collected within clinical trials) to

inform prognosis and therapy, and the ability to delineate the complex network of relationships between cancer drivers in the history and progression of the malignant process.

1. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genet.* **44,** 47–52 (2012).
2. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475,** 101–105 (2011).
3. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152,** 714–726 (2013).
4. Schuh, A. *et al.* Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120,** 4191–4196 (2012).
5. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (2014).
6. Hallek, M. *et al.* Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376,** 1164–1174 (2010).
7. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
8. Edelmann, J. *et al.* High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* **120,** 4783–4794 (2012).
9. De Paoli, L. *et al.* MGA, a suppressor of MYC, is recurrently inactivated in high risk chronic lymphocytic leukemia. *Leuk. Lymphoma* **54,** 1087–1090 (2013).
10. Schlosser, I. *et al.* Dissection of transcriptional programmes in response to serum and c-Myc in a human B-cell line. *Oncogene* **24,** 520–524 (2005).
11. Jiang, X. *et al.* Critical role of SHP2 (PTPN11) signaling in germinal center-derived lymphoma. *Haematologica* **99,** 1834–1845 (2014).
12. Zhang, J. & Chen, Q. M. Far upstream element binding protein 1: a commander of transcription, translation and beyond. *Oncogene* **32,** 2907–2916 (2013).
13. Tiacci, E. *et al.* BRAF mutations in hairy-cell leukemia. *N. Engl. J. Med.* **364,** 2305–2315 (2011).
14. Brastianos, P. K. *et al.* Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. *Nature Genet.* **46,** 161–165 (2014).
15. Heidorn, S. J. *et al.* Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell* **140,** 209–221 (2010).
16. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159,** 676–690 (2014).
17. Yang, H. *et al.* RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. *Cancer Res.* **70,** 5518–5527 (2010).
18. Jebaraj, B. M. *et al.* BRAF mutations in chronic lymphocytic leukemia. *Leuk. Lymphoma* **54,** 1177–1182 (2013).
19. Landau, D. A. & Wu, C. J. Chronic lymphocytic leukemia: molecular heterogeneity revealed by high-throughput genomics. *Genome Med.* **5,** 47 (2013).
20. Sashida, G. *et al.* ELF4/MEF activates MDM2 expression and blocks oncogene-induced p16 activation to promote transformation. *Mol. Cell. Biol.* **29,** 3687–3699 (2009).
21. Park, J. *et al.* Dyrk1A phosphorylates p53 and inhibits proliferation of embryonic neuronal cells. *J. Biol. Chem.* **285,** 31895–31906 (2010).
22. Gazda, H. T. *et al.* Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. *Am. J. Hum. Genet.* **83,** 769–780 (2008).
23. Ferreiros-Vidal, I. *et al.* Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. *Blood* **121,** 1769–1782 (2013).
24. Billot, K. *et al.* Deregulation of Aiolos expression in chronic lymphocytic leukemia is associated with epigenetic modifications. *Blood* **117,** 1917–1927 (2011).
25. Nückel, H. *et al.* The IKZF3 (Aiolos) transcription factor is highly upregulated and inversely correlated with clinical progression in chronic lymphocytic leukaemia. *Br. J. Haematol.* **144,** 268–270 (2009).
26. Beà, S. *et al.* Genetic imbalances in progressed B-cell chronic lymphocytic leukemia and transformed large-cell lymphoma (Richter's syndrome). *Am. J. Pathol.* **161,** 957–968 (2002).
27. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnol.* **30,** 413–421 (2012).
28. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122,** 3616–3627 (2013).
29. Wang, J. *et al.* Tumor evolutionary directed graphs and the history of chronic lymphocytic leukemia. *Elife* **3** (2014).
30. Stilgenbauer, S. *et al.* Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* **123,** 3247–3254 (2014).
31. Rossi, D. *et al.* Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* **121,** 1403–1412 (2013).
32. Damm, F. *et al.* Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov.* **4,** 1088–1101 (2014).
33. Winkelmann, N. *et al.* Low frequency mutations independently predict poor treatment-free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis. *Haematologica* **100,** e237–e239 (2015).
34. Puente, X. S. & Lopez-Otin, C. The evolutionary biography of chronic lymphocytic leukemia. *Nature Genet.* **45,** 229–231 (2013).
35. Nowak, M. A. *et al.* The role of chromosomal instability in tumor initiation. *Proc. Natl Acad. Sci. USA* **99,** 16226–16231 (2002).
36. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl Acad. Sci. USA* **107,** 18545–18550 (2010).
37. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155,** 948–962 (2013).
38. Hallek, M. Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. *Am. J. Hematol.* **90,** 446–460 (2015).

Author Contributions All authors contributed extensively to the work presented in this paper. D.A.L., D.N., G.G., E.T., S.S. and C.J.W. contributed to study conception and design. E.T., S.B., J.E., S.K., M.K., M.R., A.F., K.F., H.D., M.H. and S.S. performed patient selection, provided the DNA samples, and prior matched clinical and genetic data sets. C.L.S., S.G. and E.S.L. enabled sample sequencing. D.A.L., A.N.T., C.S., M.L., K.C., M.R., J.M.H., S.L.C. and G.G. contributed to the computational genomics analysis. D.A.L., E.T., J.G.R., J.B., S.K., I.B., D.M., M.A.N., D.N., G.G., S.S. and C.J.W. contributed to additional data analysis as well as manuscript preparation. All authors contributed to the writing of the manuscript.

Author Information CLL8 WES data is deposited in dbGaP under accession code phs000922.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.W. (cwu@partners.org), S.S. (Stephan.Stilgenbauer@uniklinik-ulm.de) or G.G. (gadgetz@broadinstitute.org).
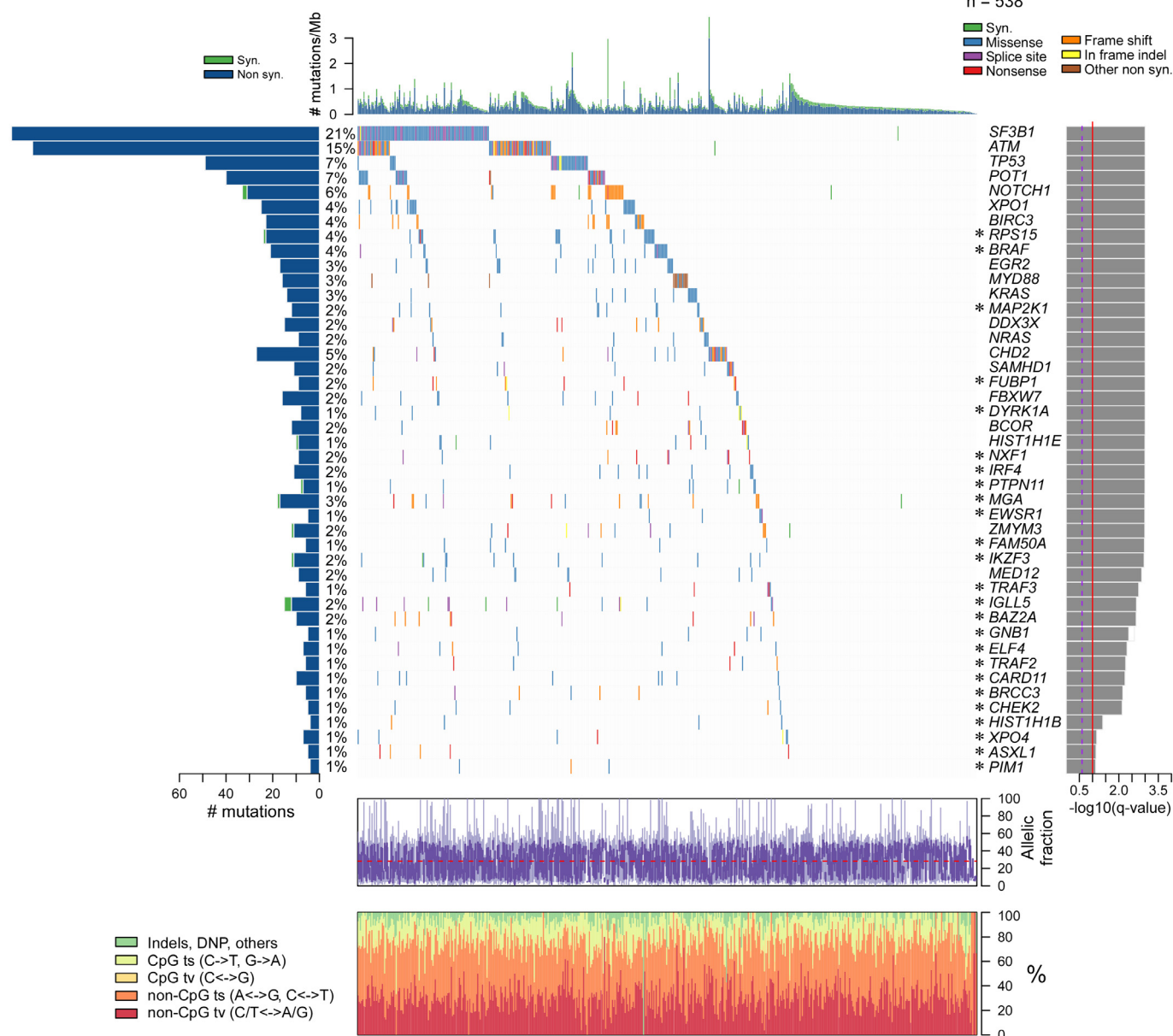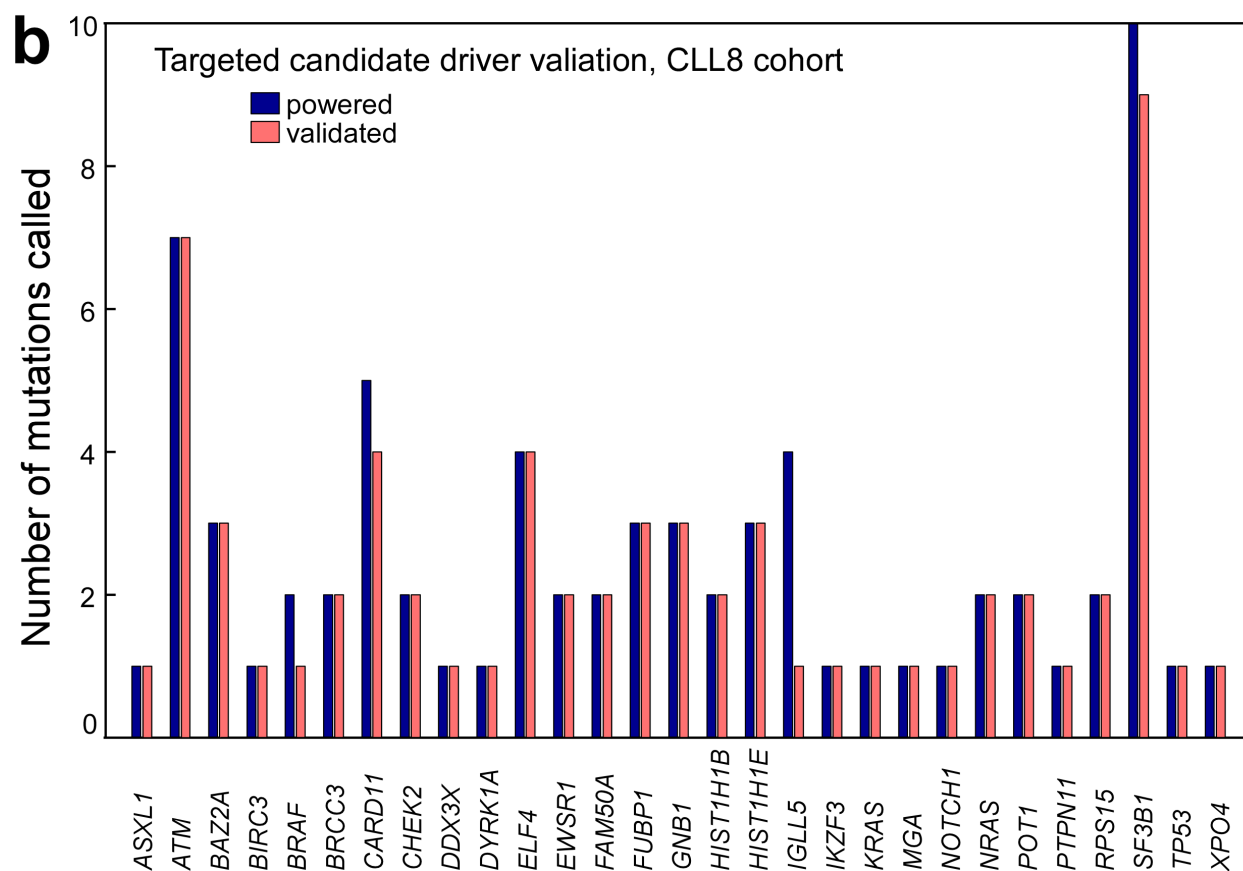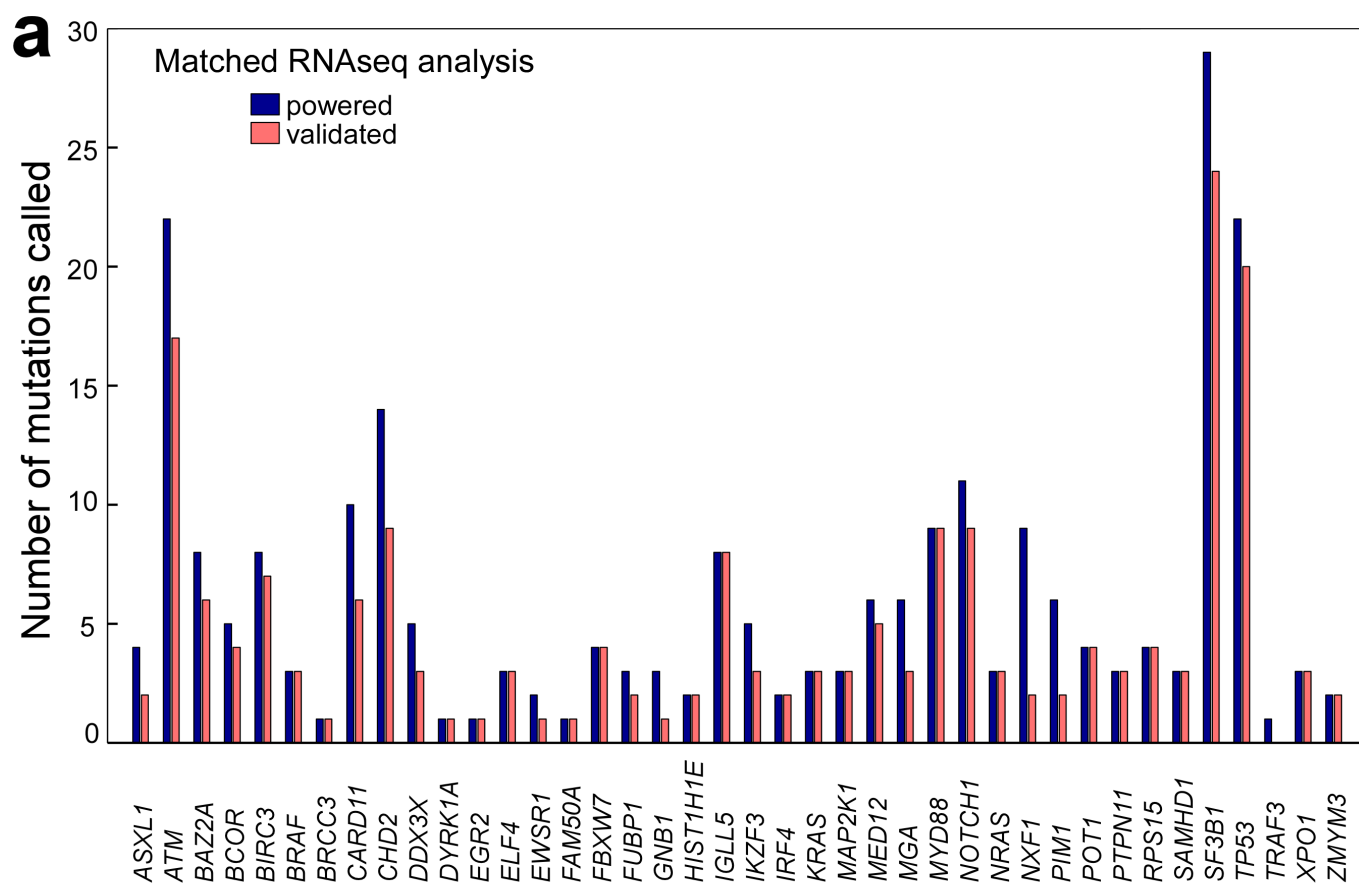
Extended Data Figure 1 | Candidate CLL cancer genes discovered in the combined cohort of 538 primary CLL samples. Significantly mutated genes identified in 538 primary CLL. Top panel: the rate of coding mutations (mutations per megabase) per sample. Centre panel: detection of individual genes found to be mutated (sSNVs or sIndels) in each of the 538 patient samples (columns), colour-coded by type of mutation. Only one mutation per gene is shown if multiple mutations from the same gene were found in a sample. Right panel: Q-values (red, Q < 0.1; purple dashed, Q < 0.25) and Hugo symbol gene identification. New candidate CLL genes are marked with asterisks. Left panel: the percentages of samples affected with mutations (sSNVs and sIndels) in each gene. Bottom panel: plots showing allelic fractions and the spectrum of mutations (sSNVs and sIndels) for each sample.
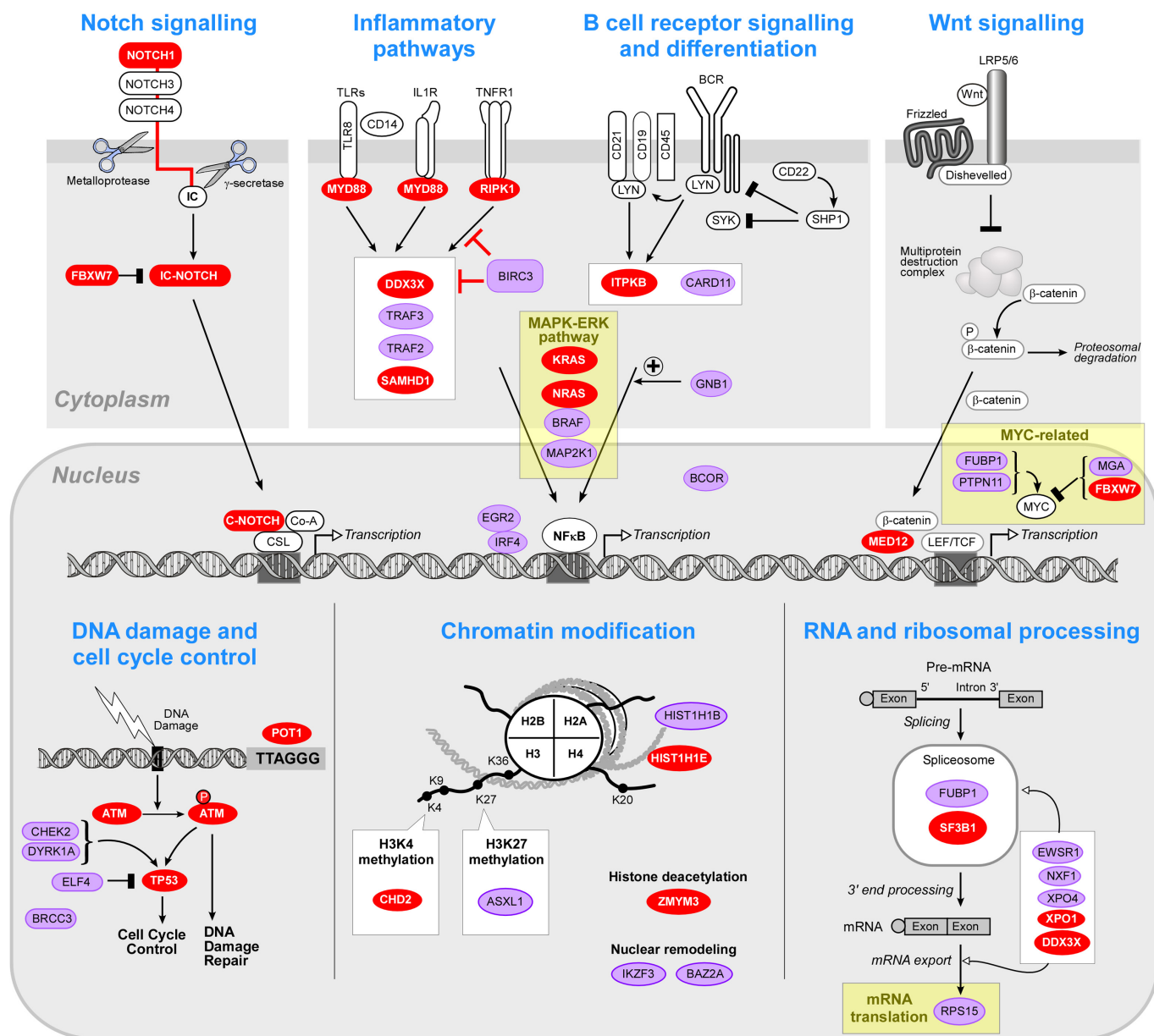
Extended Data Figure 2 | Cellular networks and processes affected by putative CLL drivers. Putative CLL cancer genes cluster in pathways that are central to CLL biology such as Notch signalling, inflammatory response and B-cell receptor signalling. In addition, proteins that participate in central cellular processes such as DNA damage repair, chromatin modification and mRNA processing, export and translation are also recurrently affected. New CLL subpathways highlighted by the current driver discovery effort are shown in yellow boxes. Red circles indicate putative driver genes previously identified[3]; purple circles indicate genes newly identified in the current study.

Extended Data Figure 3 | RNA-seq expression data for candidate CLL genes and targeted candidate driver validation. **a**, Matched RNA-seq and WES data were available for 156 CLLs (103 CLLs previously reported[3] and 53 CLLs from the ICGC studies[1]). From the WES of these 156 cases, we identified 318 driver mutations (sSNVs and sIndels). For each site, we quantified the number of alternative reads corresponding to the somatic mutation in matched RNA-seq data. We subsequently counted the number of instances in which a mutation was detected ('detected') and compared it to the number of instances in which mutation detection had >90% power based on the allelic fraction in the WES and the read depth in the RNA-seq data ('powered'). Overall, we detected 78.1% of putative CLL gene mutations at sites that had >90% power for detection in RNA-seq data. **b**, Targeted orthogonal validation (Access Array System, Fluidigm) was performed for 71 mutations (sSNVs and sIndels) in putative CLL genes, affecting 47 CLLs from the CLL8 cohort (selected on the basis of sample availability). With a mean depth of coverage of 7,472×, 65 of the 71 mutations (91.55%) validated, with a higher variant allele fraction compared with normal sample DNA (binomial $P < 0.01$).
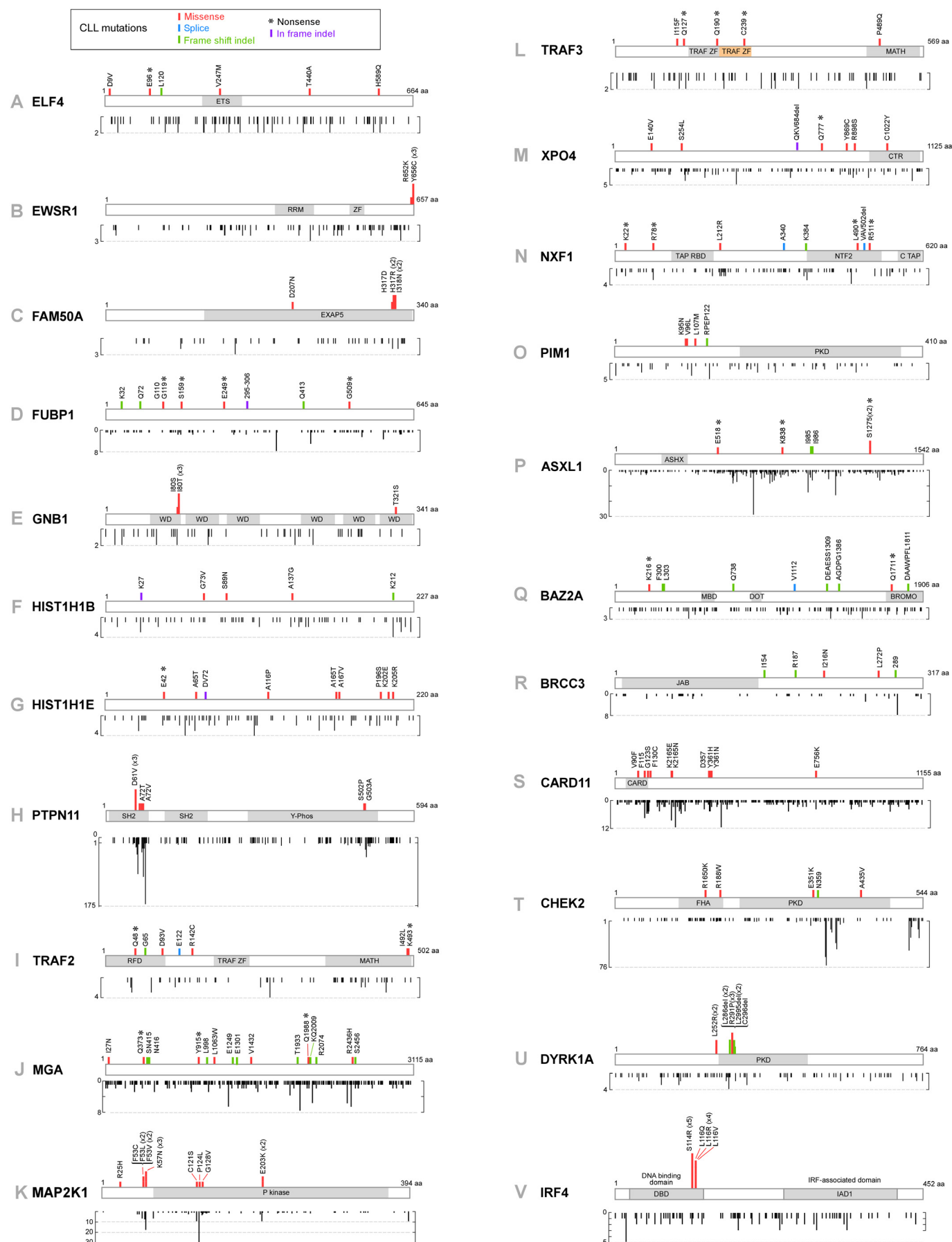
Extended Data Figure 4 | Gene mutation maps for candidate CLL genes. a–v, Individual gene mutation maps are shown for all newly identified candidate CLL cancer genes not included in Fig. 2. The plots show mutation subtype (for example, missense, nonsense) and position along the gene.
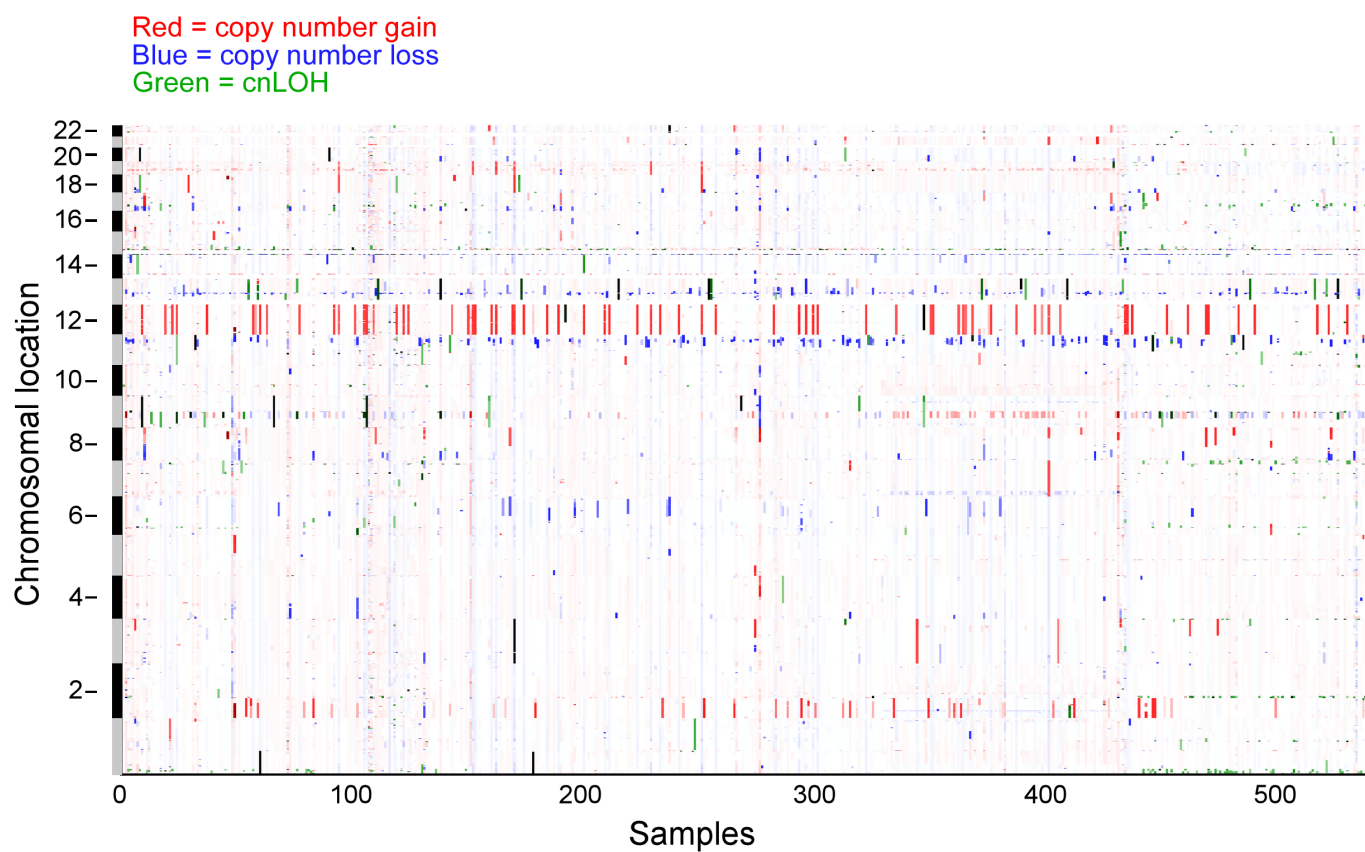
Extended Data Figure 5 | CLL copy number profiles. Copy number profile across 538 CLLs detected from WES data from primary samples (see Supplementary Methods).

Extended Data Figure 6 | Annotation of drivers based on clinical characteristics and co-occurrence patterns. **a**, Putative drivers affecting greater than 10 patients were assessed for enrichment in *IGHV* mutated versus unmutated CLL subtype (Fisher's exact test, magenta line denotes *P* = 0.05). **b**, Putative drivers affecting greater than 10 patients were assessed for enrichment in samples that received therapy before sampling (Fisher's exact test). Putative drivers affecting greater than 10 patients were tested for co-occurrence. **c**, **d**, Significantly high (**c**) or low (**d**) co-occurrences are shown (*Q* < 0.1, Fisher's exact test with Benjamini Hochberg, false discovery rate, after accounting for prior therapy and *IGHV* mutation status, see Supplementary Methods).
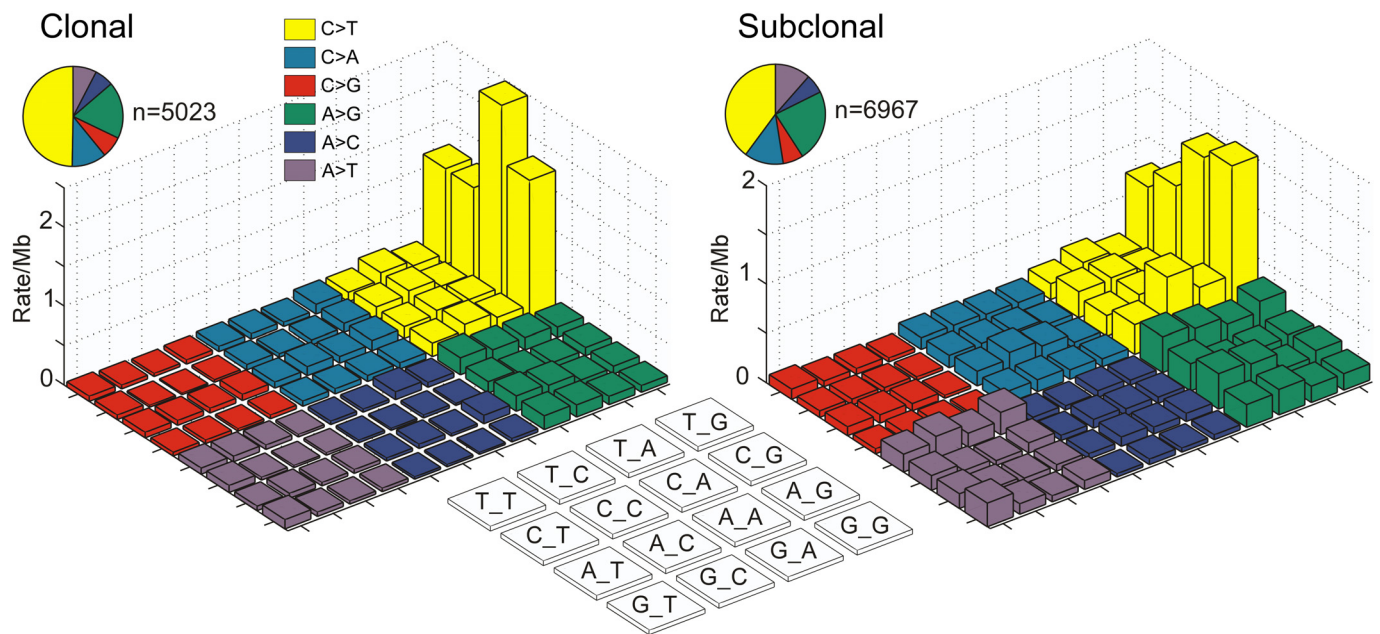
Extended Data Figure 7 | Mutation spectrum analysis, clonal versus subclonal sSNVs. The spectrum of mutation is shown for the clonal and subclonal subsets of coding somatic sSNVs across WES of 538 samples. The rate is calculated by dividing the number of trinucleotides with the specified sSNVs by the covered territory containing the specified trinucleotide. Both clonal and subclonal sSNVs were similarly dominated by C > T

transitions at CpG sites. Thus, this mutational process that was previously associated with ageing[39] not only predates oncogenic transformation (since clonal mutations will be highly enriched in mutations that precede the malignant transformation[40]) but also is the dominant mechanism of malignant diversification after transformation in CLL.

39. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
40. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339,** 1546–1558 (2013).
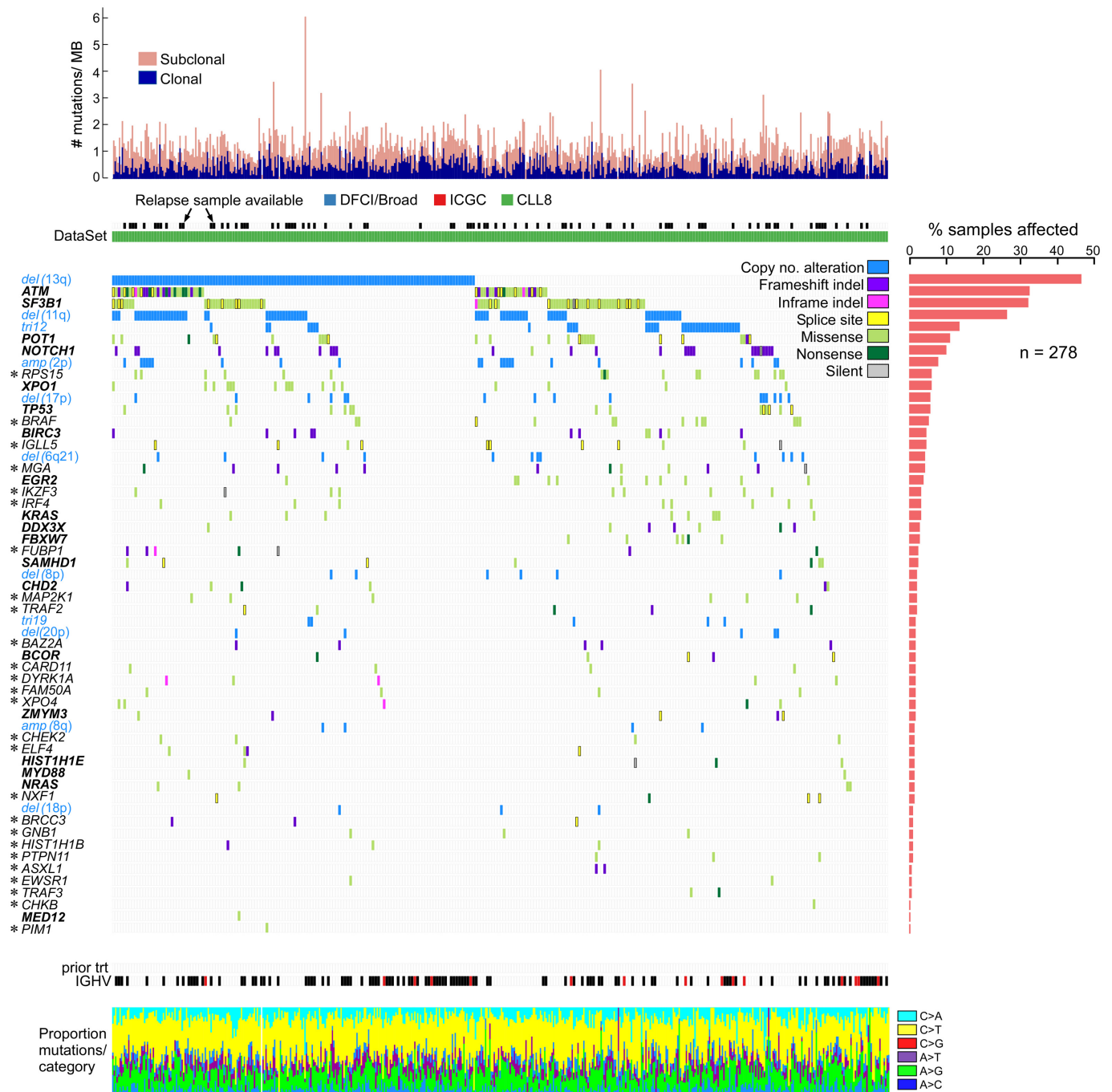
Extended Data Figure 8 | The CLL driver landscape in the CLL8 cohort.
Somatic mutation information shown across the 55 candidate CLL cancer genes and recurrent somatic CNAs (rows) for 278 CLL samples collected from patients enrolled on the CLL8 clinical trial primary that underwent WES (columns). Recurrent somatic CNA labels are listed in blue, candidate CLL cancer genes are listed in bold if previously identified in Landau et al.[3], and with an asterisk if newly identified in the current study.
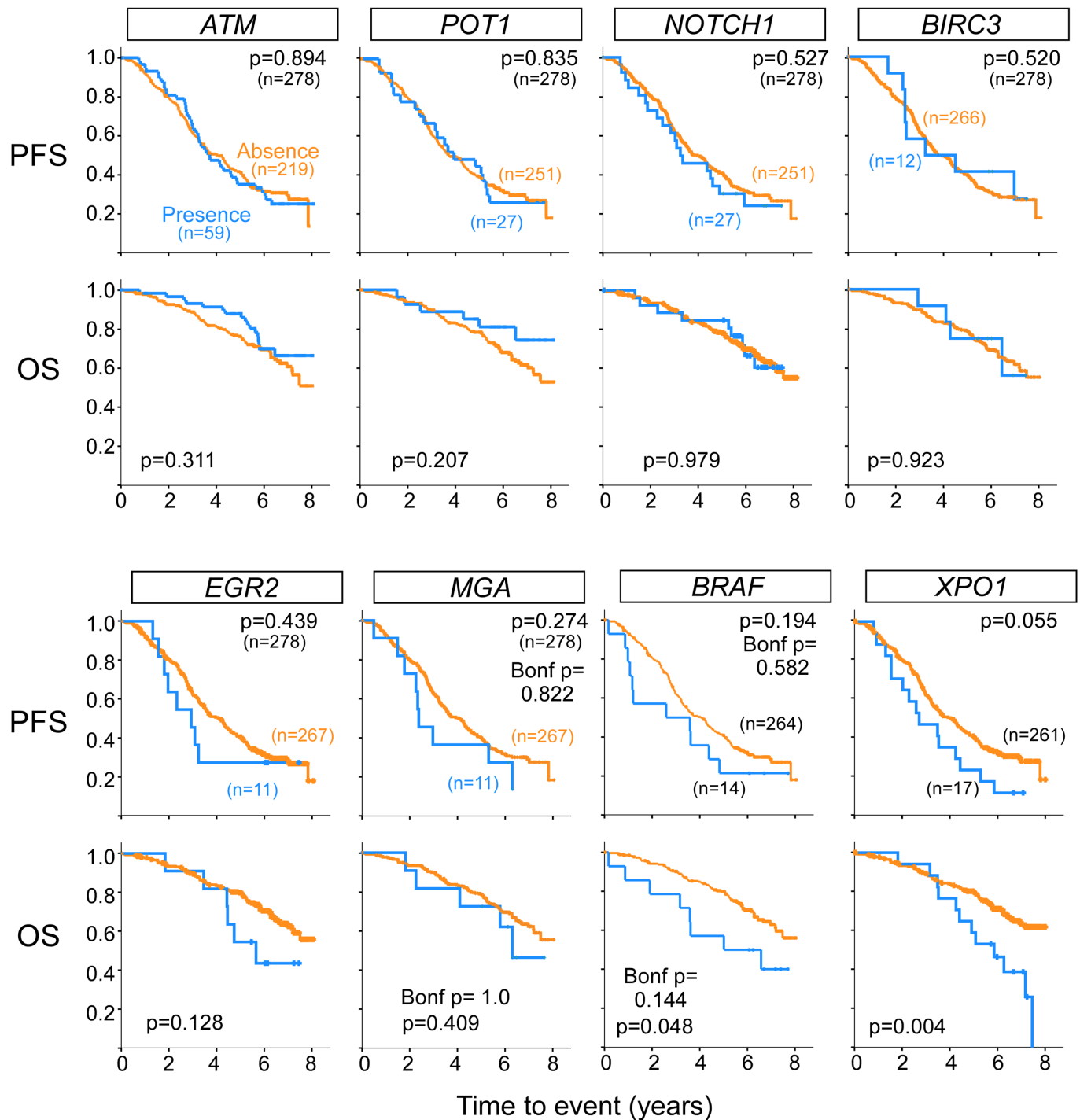
Extended Data Figure 9 | CLL8 patient cohort clinical outcome (from 278 patients) information by CLL cancer gene. Kaplan–Meier analysis (with logrank *P* values) for putative drivers not associated with significant impact on progression-free survival (PFS) or overall survival (OS) in the cohort of 278 patients that were treated as part of the CLL8 trial. For candidate CLL genes tested here for the first time regarding impact on outcome, a Bonferroni *P* value is also shown.
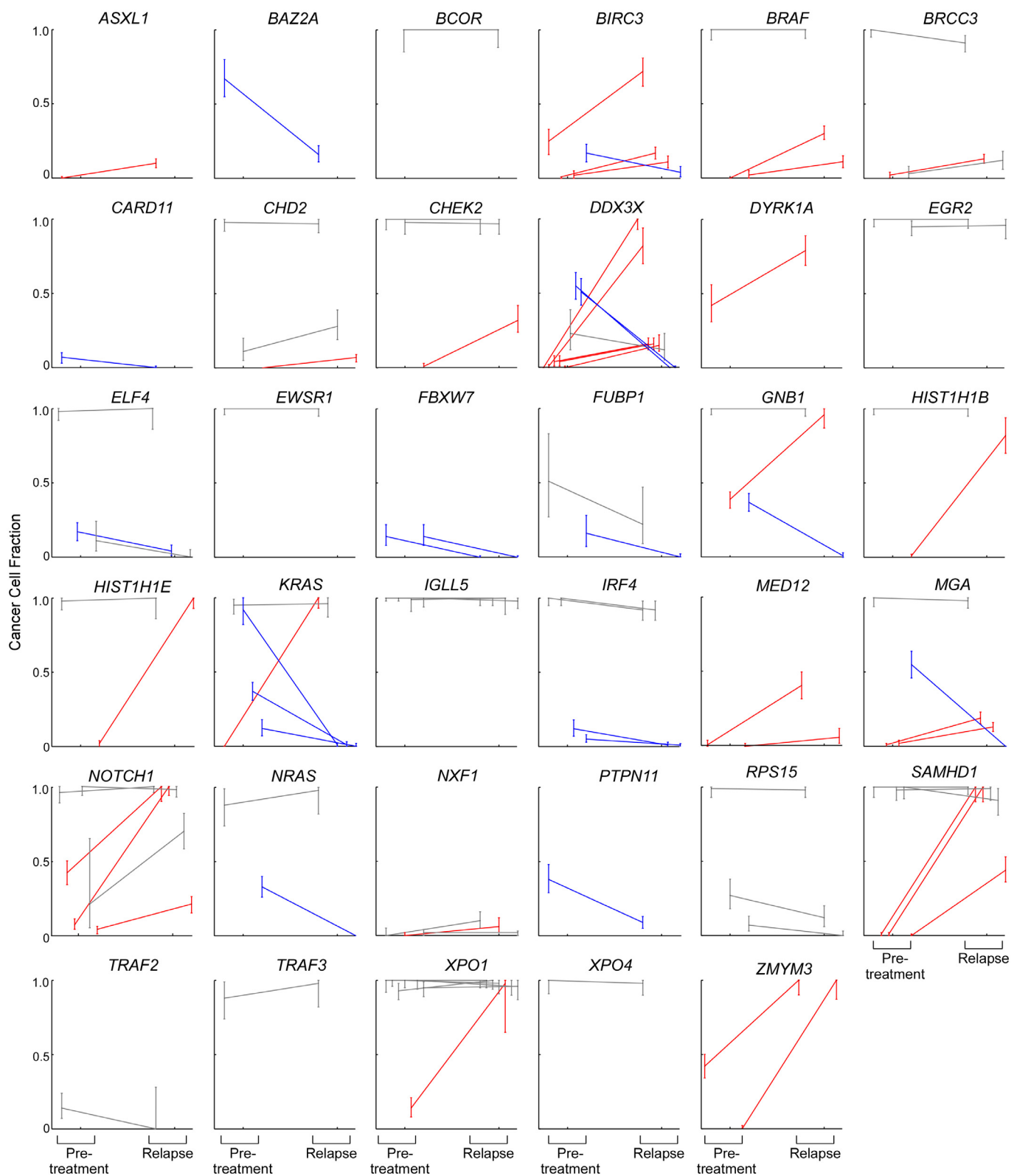
Extended Data Figure 10 | Comparison of pre-treatment and relapse cancer cell fraction (CCF) for non-silent mutations in candidate CLL genes across 59 CLLs. For each CLL gene mutated across the 59 CLLs that were sampled longitudinally, the modal CCF is compared between the pre-treatment and relapse samples. CCF increases (red), decreases (blue) or stable CCF (grey) over time are shown (in addition to CLL genes shown in Fig. 5). A significant change in CCF over time (red or blue) was determined if the 95% CI of the CCF in the pre-treatment and relapse samples did not overlap.

Single cell activity reveals direct electron transfer in methanotrophic consortia

Shawn E. McGlynn[1]†*, Grayson L. Chadwick[1]*, Christopher P. Kempes[2,3,4] & Victoria J. Orphan[1]

Multicellular assemblages of microorganisms are ubiquitous in nature, and the proximity afforded by aggregation is thought to permit intercellular metabolic coupling that can accommodate otherwise unfavourable reactions. Consortia of methane-oxidizing archaea and sulphate-reducing bacteria are a well-known environmental example of microbial co-aggregation; however, the coupling mechanisms between these paired organisms is not well understood, despite the attention given them because of the global significance of anaerobic methane oxidation. Here we examined the influence of interspecies spatial positioning as it relates to biosynthetic activity within structurally diverse uncultured methane-oxidizing consortia by measuring stable isotope incorporation for individual archaeal and bacterial cells to constrain their potential metabolic interactions. In contrast to conventional models of syntrophy based on the passage of molecular intermediates, cellular activities were found to be independent of both species intermixing and distance between syntrophic partners within consortia. A generalized model of electric conductivity between co-associated archaea and bacteria best fit the empirical data. Combined with the detection of large multi-haem cytochromes in the genomes of methanotrophic archaea and the demonstration of redox-dependent staining of the matrix between cells in consortia, these results provide evidence for syntrophic coupling through direct electron transfer.

Ecological processes are fundamentally spatial in nature: those governing microbial organisms are no exception. The ubiquity and impact of biofilms, consortia, and other multicellular assemblages in the fields of environmental microbiology, industry, and medicine demonstrates the necessity of relating the spatial position of cells to metabolic activity and community function[1–7]. Theoretical modelling[8] and laboratory experiments with artificial co-cultures[4,9] have offered fundamental insights regarding the effect of spatial architecture on fitness and physiology of interacting populations, but studying the influence of spatial organization on uncultured microorganisms has remained a long-standing challenge. To translate information learned from modelling and derived laboratory results to systems found in nature requires new methodological strategies that are capable of elucidating microbial structure–activity relationships.

Here, fluorescence *in situ* hybridization and nanoscale secondary ion mass spectrometry (FISH–nanoSIMS) combined with ^{15}N stable isotope probing was used to investigate how single-cell metabolic activity is related to cellular configuration in highly structured, bi-species microbial consortia in environmental samples (Fig. 1 and Extended Data Fig. 1). We applied these methods to empirically test long-standing hypotheses regarding the metabolic interactions underpinning the environmentally important microbial symbiosis responsible for the anaerobic oxidation of methane (AOM) in ocean sediments[10–13]. Discovered over a decade ago[10], these consortia consist of multiple lineages of as yet uncultured anaerobic methanotrophic archaea (ANME) and sulfate-reducing Deltaproteobacteria (SRB), and form diverse aggregate configurations within methane seep sediments worldwide[14,15]. Initial FISH–SIMS acquired whole aggregate stable isotope depth profiles offered isotopic evidence for the involvement of ANME–SRB consortia in anaerobic methanotrophy[11] and documented broad anabolic activity patterns amongst different AOM aggregate morphologies[16], but key questions regarding the

mechanism of this syntrophic association remain. There are a number of hypotheses regarding the metabolic interactions underlying this enigmatic methane-fueled symbiosis, ranging from classical syntrophy based on hydrogen, formate or acetate[17,18], to less conventional forms of metabolite or reducing equivalent exchange (for example, methanethiol and disulfide)[19,20]. Understanding whether there is a universal mechanism controlling ANME–SRB mediated methane oxidation or if different archaeal–bacterial AOM consortia use a variety of syntrophic strategies, is still an unresolved question in the field.

Independent of the specific mechanism, a key prediction regarding syntrophic associations of microbes is that the spatial arrangement of paired organisms can greatly influence the metabolic activity of individual cells[12,13,21,22]. In these cases, homogeneous species mixing is expected to facilitate efficient transfer of diffusible intermediates and lead to enhanced metabolic activity[13,22]. Similarly, at the single-cell level, syntrophic partners in immediate proximity to one another are expected to gain a greater metabolic benefit in comparison to cells that lack a syntrophic interface. For the AOM system, these predictions are captured in previously published models[12,13] and also in examples presented here (Extended Data Fig. 2) which are based on syntrophic transfer[23] or commensal passage[20] of a diffusible intermediate. Notably, these modelling results are at odds with the frequent documentation of large environmental consortia with spatially segregated ANME and SRB cells and it remains unclear how these configurationally segregated consortia persist, and often dominate, if they are at a disadvantage to those that are well mixed[10,16]. The discrepancies between *in situ* observations and model predictions motivate a series of hypotheses that are testable with single-cell biosynthetic activity measurements of AOM consortia: (1) within highly segregated consortia, the vast majority of activity will be restricted to cells at interfaces between syntrophic partners, and (2) that segregated syntrophic consortia will have lower total activity levels on average

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. [2]Exobiology Branch, National Aeronautics and Space Administration Ames Research Center, Moffett Field, California 94035, USA. [3]Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA. [4]SETI Institute, Mountain View, California 94034, USA. †Present address: Department of Biological Sciences, Tokyo Metropolitan University, Tokyo 192-0397, Japan. *These authors contributed equally to this work.
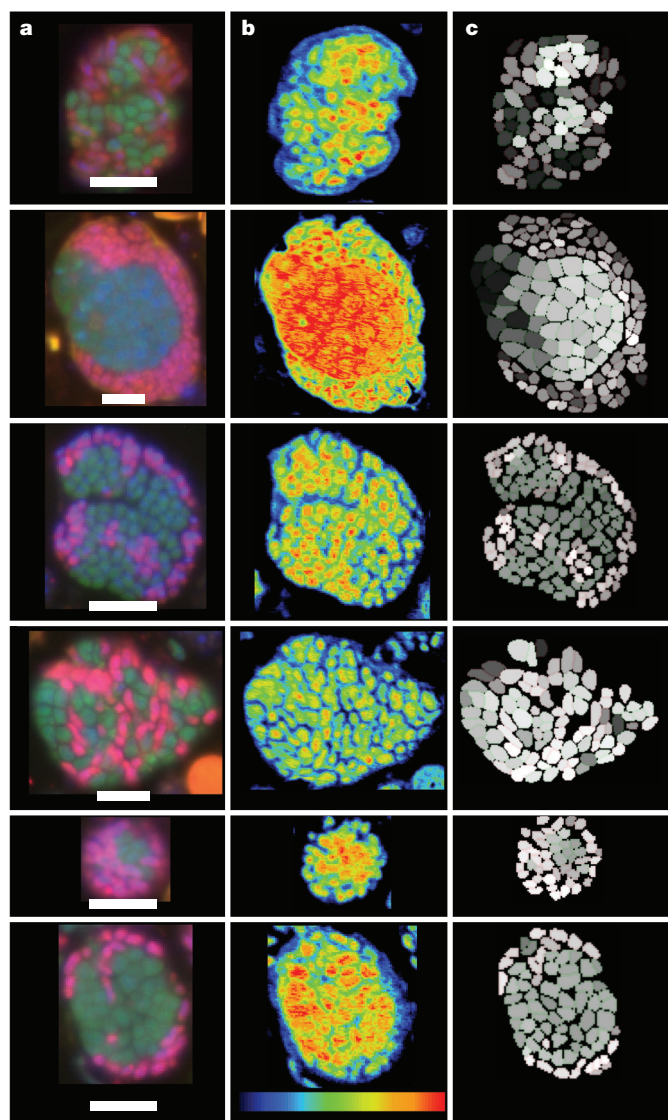
Figure 1 | Examples of AOM consortia identified by FISH and paired anabolic activity measurement via nanoSIMS. a, FISH-identified consortia showing archaeal cells (green) and Deltaproteobacteria (pink). The top two panels represent consortia of ANME-2c or 2b paired with Deltaproteobacteria. The lower four panels show ANME-2c archaea paired with the seep-specific deltaproteobacterial group, SEEP-SRB1a. Scale bars, 3 µm. **b**, Corresponding nanoSIMS ion images of biomass show $^{14}N^{12}C^{-}$ ion images with warmer colours indicating higher secondary ion counts (maximum 1,500 counts). **c**, Single cell activities are measured as ^{15}N atom percentages for regions of interest (ROI) representing the FISH-identified archaea and bacteria in each consortium. Lighter shaded cells are more enriched in ^{15}N, which corresponds with higher levels of anabolic activity and $^{15}NH_4^+$ assimilation. Representative aggregates were chosen from the larger data set composed of 62 aggregates.

relative to well-mixed consortia. If these patterns of activity are not observed, then the interactions driving the symbiosis may be distinct from the classical view of syntrophy occurring through the exchange of a diffusible chemical intermediate.

These hypotheses were evaluated for phylogenetically diverse ANME-2 archaea (belonging to the order Methanosarcinales) and partner Deltaproteobacteria using high-resolution biosynthetic activity measurements paired with FISH-based microbial identification; giving us the ability to catalogue cell activity, phylogeny, and cellular position within consortia. ^{15}N-ammonium assimilation, used as a marker for biosynthetic activity[16,24], was determined for 5,453 FISH-identified cells within 62 consortia from a deep-sea sediment incubation showing AOM activity, allowing an assessment of the biosynthetic
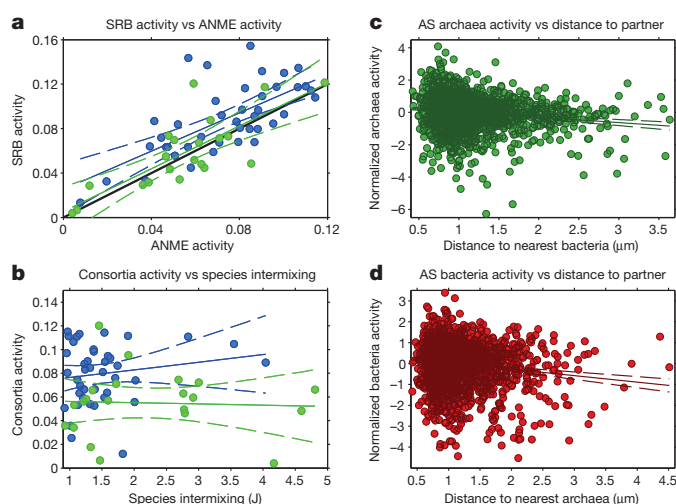
Figure 2 | Activity relationships between archaea and bacteria in AOM consortia. a, Population-level average of bacterial activity versus archaeal activity for individual AOM consortia revealing a positive correlation in activity between paired partners. Individual AS aggregates shown in blue ($n = 41$, $R^2 = 0.47$), AD in green ($n = 21$, $R^2 = 0.62$). The 1:1 line is shown in black. **b**, Total consortium activity plotted against the join index (J) of spatial intermixing between archaeal and bacterial partners, where lower J values represent greater mixing (AS in blue: $n = 41$, $R^2 = 0.03$; AD in green: $n = 21$, $R^2 = 0.00$). **c, d**, Activity of single archaeal or bacterial cells, respectively, plotted against distance to the nearest syntrophic partner for the AS data set (**c**, $n = 1,967$, $R^2 = 0.02$; **d**, $n = 2,067$, $R^2 = 0.02$). The equivalent analysis for AD is provided in Extended Data Fig. 6. In all plots R^2 values and solid lines represent linear regressions of the plotted data. Dashed lines illustrate the 95% confidence intervals in slopes and intercepts of the linear regressions.

activity of each cell as it relates to spatial positioning of adjacent syntrophic partners (Fig. 1). Two different groups of co-occurring AOM consortia in the incubation were analysed: 41 archaeal aggregates paired with the specific Desulfobacteraceae lineage SEEP-SRB1a which have been observed as a common bacterial partner of ANME-2 world-wide[25] (AS; ANME-2c: SEEP-SRB1a), and 21 archaeal aggregates paired with other, non-SEEP-SRB1a Deltaproteobacteria (AD; ANME-2c or ANME-2b: Deltaproteobacteria); (Fig. 1, Extended Data Fig. 3 and Supplementary Information).

Spatial patterns of cellular activity

FISH analyses of AS and AD aggregates revealed nearly equal abundances of archaea and bacterial cells within each consortia (Extended Data Fig. 3). The average activity (^{15}N enrichment) of archaeal and bacterial populations from each consortia was found to be correlated and close to the 1:1 line suggesting a beneficial metabolic interaction (Fig. 2a), although the AS bacteria were on average slightly more active than their archaeal partners (Extended Data Fig. 4). The AS and AD consortia analysed in this study occupied a considerable range of both biosynthetic activity and cell number (Extended Data Fig. 3). In order to relate this range of activity to the amount of partner intermixing, we developed a quantitative metric, J, to describe aggregate spatial mixing (Extended Data Fig. 5 and Supplementary Discussion). J values for aggregates represented a range of mixing, and permutation tests revealed that the majority of AS and AD consortia exist in conformations where partnering cells were more segregated than could be explained by random chance, consistent with patterns emerging from binary cell division (Extended Data Fig. 5). Across this conformational variability however, the degree of mixing between cell types (J values between 0.9 and 4.8) did not influence the average biosynthetic activity of the entire consortium (Fig. 2b), indicating the overall activity of well-mixed consortia (low J values) was not greater than those that were segregated.

When examining the ^{15}N-based activity patterns of individual archaeal and bacterial cells within a consortium, each cell's activity

was found to be unrelated to proximity to the nearest archaeal or bacterial partner (Fig. 2c, d and Extended Data Fig. 6). Similarly, in the vast majority of consortia, the cells located at syntrophic interfaces were not significantly more active than those surrounded by the same cell type (Supplementary Tables 1–4, and Extended Data Fig. 7). These single cell measurements are consistent with previously published results of intact AOM consortia, where a correlation between bulk ^{15}N enrichment and natural abundance $\delta^{13}C$ (a proxy for methanotrophic ANME biomass) in SIMS depth profiles was not detected[16]. We also examined whether archaeal or bacterial cellular activity was related to the external environment, however no significant correlation with distance to the aggregate–environment interface was observed (Extended Data Fig. 8). These observations are contrary to conventional diffusive model results for syntrophic AOM partners[12,13] in scenarios where one or both partners are dependent on the surrounding environment (Extended Data Fig. 2 shows diffusion model results for a broad range of diffusion and activity rates). Other potential relationships between total consortia activity and aggregate size approximated from total cell number, as well as the relative ratio of ANME and SRB cells in each consortium were also not significant in our data set (R^2 values <0.04).

Diffusion versus direct electron transfer

The distance-independent trends in cellular activity presented in Fig. 2b–d, Extended Data Fig. 6, and Supplementary Tables 1–4 are in stark contrast to what is predicted in the case of syntrophic exchange[12,13] or commensal sharing[20] of a diffusible intermediate (see diffusion model results in Extended Data Fig. 2 and Supplementary Information). To explore diffusion-independent scenarios which might explain our empirical data, we constructed a second generalized model that captures the basic features of direct interspecies electron transfer[26]. This model is based on electron export by one cell type and electron import by a partner within the consortia, where electrons are able to freely flow across the entire aggregate with a dependence on electric potential (Supplementary Information). Consistent with our data, these models predict a reduction in the overall correlation between biosynthetic activity and aggregate geometry (J metric), especially as the electric conductivity of the aggregate is increased relative to the growth rates (Extended Data Fig. 9). In particular, total aggregate activity is relatively insensitive to how well mixed the aggregate is, and single-cell activities are less correlated with the distance to either the aggregate surface or the syntrophic partner (Extended Data Fig. 9). In both models (Extended Data Figs 2 and 9), the empirical results presented in Fig. 2 are best matched with an increased ratio of metabolic exchange rates relative to cellular activity rates and both models converge to similar results for high rates of transport. For a diffusible intermediate, this ratio would need to be larger than predictions from known intermediate diffusivities[13], our observed growth rates, and the expected growth yields.

These model results related to the J metric illustrate key geometric and mechanistic differences beyond those stemming from spatial gradients. We find that in the low transport regimes for both the diffusive and electric conductivity models, the total activity of consortia is strongly related to the overall mixing between the two partners, but with opposite and somewhat unanticipated outcomes. As predicted, the slow relative transport diffusion scenario shows the highest activity associated with well-mixed consortia (low J values), however, the electron conductivity model indicates higher levels of activity in more segregated consortia (high J values); (Extended Data Fig. 9a). This prediction arises because our conductive treatment of the consortia relies on the global electric potential for each consortia, which is strongest when the electron producing and consuming cells are spatially segregated, maximizing polar charge separation. It should also be noted that mechanisms of electron diffusion[27,28] would produce relative transport rates sufficient for matching the observed equable activity patterns (Supplementary Discussion).

Multi-haem cytochrome genes in ANME-2 genomes

Motivated by these modelling results in which direct electron transfer apparently relaxes spatial controls on aggregate activity in agreement with our single cell observations, we analysed available ANME-2 genomes to determine whether there were signs of this alternative mode of syntrophy as has been previously suggested for ANME-1 (ref. 29). Remarkably, the genomes of two recently sequenced methanotrophic archaea (ANME-2a[30] and ANME-2d[31]), as well as a reconstructed metagenomic bin corresponding to the ANME-2b (data not shown) were each found to encode large multi-haem cytochromes (MHCs), including the largest described from an archaeon to date (34 haems); (Fig. 3a, b). A subset of these previously overlooked MHCs occur fused with a single putative S-layer domain which appears to be a homologue of the S-layer protein in *Methanosarcina acetivorans* (Fig. 3a), suggesting MHC export from the cell and incorporation into the archaeal S-layer. The occurrence of MHCs of this size encoded in a genome is rare even in bacteria, and those that do occur are almost exclusively found in organisms known to conduct extracellular electron transfer such as *Geobacter* and *Shewanella* — species which serve as model organisms for the process[26,32–34].

Cytochrome reactive staining in consortia

To test possibility that MHCs are positioned between cells within consortia as electron conduits, the cytochrome reactive histochemical stain 3,3′-diaminobenzidine (DAB)[35] was applied to AOM consortia recovered directly from sediment. Treatment with DAB and H_2O_2 followed by post fixation with OsO_4 resulted in the staining of: (1) the cellular membranes of both syntrophic partners, (2) some intra-cellular membrane invaginations of paired Deltaproteobacteria, and (3) the extracellular space between cells within consortia (Fig. 3c, d). DAB staining with the addition of H_2O_2 was observed in many, but not all aggregates in the preparation, suggesting the possibility of phylogenetic or phenotypic variation in extracellular MHC production within the sediment-hosted AOM consortia. No visible staining was observed in control experiments without H_2O_2 (Fig. 3e, f). As DAB is known to react with redox active transition metal ions (including those bound by haem groups within cytochromes) in the presence of H_2O_2[36], these results are consistent with the localization of the respiratory chain in the cellular membrane for each organism, and also with the presence of haem proteins capable of redox activity in the space between cells in consortia.

A model for direct electron transfer

The electron microscopy results reported above, together with the presence of the large MHCs in all available ANME-2 genomes suggests that extracellular electron transfer may be an important feature of the anaerobic methanotroph lifestyle. Based on this finding, and the lack of genomic evidence presented for other syntrophic models[20], we propose the catabolic model for AOM coupled to extracellular electron transfer depicted in Fig. 4. Using known biochemical coupling mechanisms in methanogens[37], the oxidation of one mole of methane can result in four moles of reduced methanophenazine. We propose that these methanophenazines are oxidized by an integral membrane protein (for example, cytochrome *b*), with electrons being transferred onto an initial MHC for transport from the membrane to the S-layer. Tandem proteins 2566125773 and 2566125774 in the ANME-2a genome encode a predicted formate dehydrogenase-related cytochrome *b* and an 11-haem multi-haem cytochrome, which are possible candidates for this methanophenazine:cytochrome *c* oxidoreductase step. The MHC/S-layer fusion proteins depicted in Fig. 3a could then be used for electron transport across the S-layer. Finally, large extracellular cytochromes such as the 31 CxxCH motif containing protein 2566123495 and numerous other small MHC proteins could be used to confer electrical conductivity to the exopolymer matrix between the ANME-2 archaea and their SRB partners, similar to the case of MHC proteins thought to facilitate growth of thick geobacter biofilms[26]. As formulated here, this proposed metabolic pathway could potentially
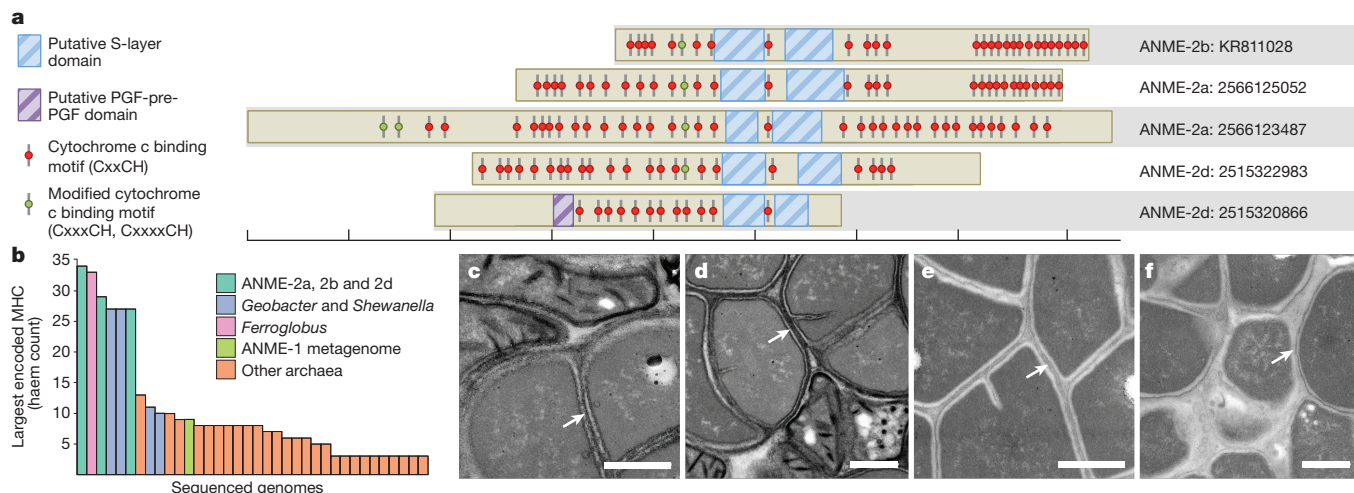
Figure 3 | Multi-haem cytochrome genes, genomes and TEM visualization of haem group reactivity in representative ANME–SRB consortia. **a**, Five MHCs with predicted S-layer domains from reconstructed ANME-2 genomes. Cytochrome *c* binding motif sites (CxxCH) are indicated by red dots, alternative binding motifs (CxxxCH and CxxxxCH) are shown in green. Putative S-layer domains and potential PGF-pre-PGF archaeosortase recognition domains as predicted by the NCBI conserved domain database are shown in blue and purple, respectively. Protein schematics are simply positioned to show compositional similarity, not sequence alignment. Each vertical tick-mark denotes 250 amino acids. The gene identifier numbers are shown on the right. **b**, Size of largest MHC present in sequenced archaeal genomes with representative model bacteria, *Geobacter* and *Shewanella*, included as a reference. **c–f**, Transmission electron microscopy (TEM) micrographs of sediment-hosted methanotrophic consortia treated with the haem-reactive compound 3′-3-diaminobenzadine (DAB). **c, d**, Positive staining of the membranes and extracellular space between archaeal and bacterial cells in the presence of H_2O_2. **e, f**, Control cells from DAB experiments where H_2O_2 was omitted. Scale bars, 500 nm. Arrows mark the interfaces of cells.

result in the net translocation of $\sim 2H^+$ per CH_4 oxidized by the ANME-2 archaea, with some uncertainty due to the exact stoichiometry of the proton and sodium pumping complexes involved. This low proton efflux per substrate used fits well with the small thermodynamic free energies associated with anaerobic oxidation of methane[23] and the slow growth rates of these organisms.

Together, the evidence from our spatially resolved analysis of cellular activity, genomic observations, and electron microscopy experiments is highly suggestive of direct interspecies electron transfer between methanotrophic ANME-2 and associated Deltaproteobacteria. This diffusion-independent mechanism appears to largely obviate the geometric constraints amongst ANME–SRB consortia. Interspecies electron transfer may also contribute to greater stability of the association compared with syntrophic exchange of a diffusible intermediate, where loss to the environment and greater sensitivity to environmental chemical fluctuations can limit otherwise favourable thermodynamics. The type of interspecies electronic coupling described here and in co-cultured organisms for example[38,39], (and see also ref. 26 and references therein), may be an underappreciated natural phenomenon that contributes to microbial niche construction, where metabolic coupling facilitated by direct electron transfer could function as a means of generating stable syntrophic microbial assemblages. Additionally, the MHCs in ANME-2 genomes may help explain the occurrence of ANME-2 aggregates without syntrophic partners, as well as the observation of AOM with metal oxides, where the ANME-2 may be able to grow on their own as metal oxide reducers[40]. Future work will be required to fully comprehend the detailed mechanisms of electron transfer, the role of MHCs, and the potential function of interspecies electron transfer among different ANME groups or habitats. The culture-independent approach described here is applicable to investigating interactions occurring in a broad range of environmental microbial assemblages and may be amenable with other stable isotope tracers (for example, deuterated water), where the central challenge is to understand how metabolic interconnectivity and spatial relationships between organisms drives local and bulk geochemical processes.

Note added in proof: While this manuscript was in review, a paper noting the presence of large multi-haem cytochromes in archaea, including the ANME-2d genome, was published[41].

Figure 4 | A proposal for the energy metabolism of ANME-2a. The syntrophic half-reaction for ANME-2a is $CH_4 + 2H_2O \rightarrow CO_2 + 8e^- + 8H^+$. Reducing equivalents from methane are generated through the methyl branch of the Wood–Ljungdahl pathway and are deposited in the membrane-bound methanophenazine pool (Mp/MpH₂) via reactions at the Hdr, Fpo, and Rnf complexes, which oxidize CoM-SH/CoB-SH, $F_{420}H_2$, and Fd_{red}, respectively. All proteins involved in the reverse methanogenesis pathway (green) have been identified in the ANME-2a genome[27]. We have identified proteins (orange) in the ANME-2 genomes that may be responsible for extracellular electron transport as outlined in the text. Haem groups are shown schematically as in Fig. 3, and the number of haems predicted to be bound by the peptide (CxxCH motifs) is indicted.

1. Tolker-Nielsen, T. & Molin, S. Spatial organization of microbial biofilm communities. *Microb. Ecol.* **40,** 75–84 (2000).
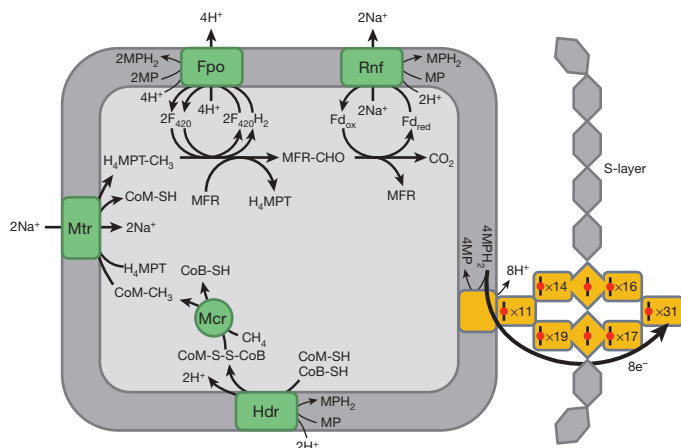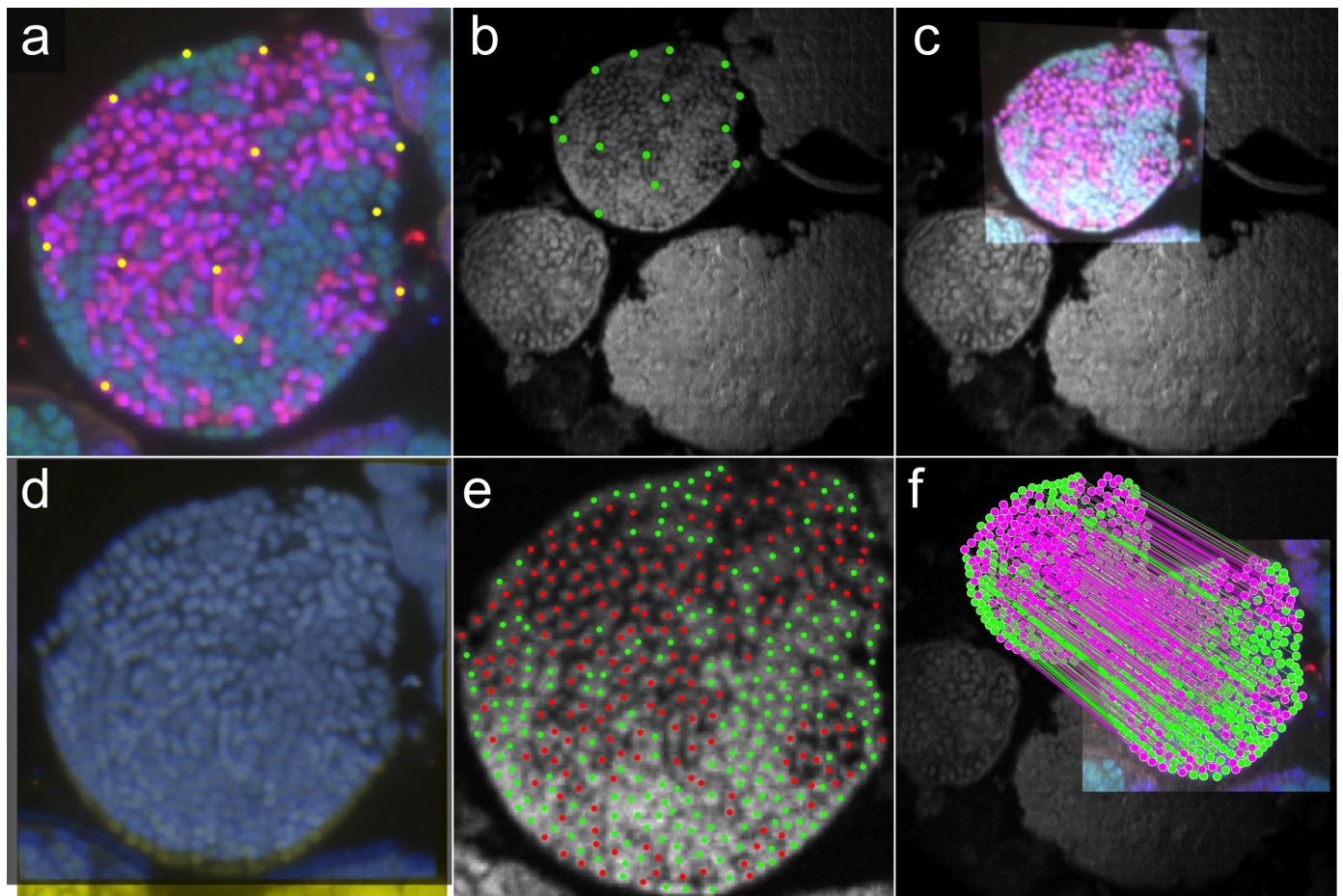
2. Rickard, A. H., Gilbert, P., High, N. J., Kolenbrander, P. E. & Handley, P. S. Bacterial coaggregation: an integral process in the development of multi-species biofilms. *Trends Microbiol.* **11,** 94–100 (2003).

3. Battin, T. J. *et al.* Microbial landscapes: new paths to biofilm research. *Nature Rev. Microbiol.* **5,** 76–81 (2007).

4. Kim, H. J., Boedicker, J. Q., Choi, J. W. & Ismagilov, R. F. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proc. Natl Acad. Sci. USA* **105,** 18188–18193 (2008).

5. Wintermute, E. H. & Silver, P. A. Dynamics in the mixed microbial concourse. *Genes Dev.* **24,** 2603–2614 (2010).

6. Wessel, A. K., Hmelo, L., Parsek, M. R. & Whiteley, M. Going local: technologies for exploring bacterial microenvironments. *Nature Rev. Microbiol.* **11,** 337–348 (2013).

7. Momeni, B., Brileya, K. A., Fields, M. W. & Shou, W. Strong inter-population cooperation leads to partner intermixing in microbial communities. *eLife* **2,** e00230 (2013).

8. Kempes, C. P., Okegbe, C., Mears-Clarke, Z., Follows, M. J. & Dietrich, L. E. P. Morphological optimization for access to dual oxidants in biofilms. *Proc. Natl Acad. Sci. USA* **111,** 208–213 (2014).

9. Nielsen, A. T., Tolker-Nielsen, T., Barken, K. B. & Molin, S. Role of commensal relationships on the spatial structure of a surface-attached microbial consortium. *Environ. Microbiol.* **2,** 59–68 (2000).

10. Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407,** 623–626 (2000).

11. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293,** 484–487 (2001).

12. Orcutt, B. & Meile, C. Constraints on mechanisms and rates of anaerobic oxidation of methane by microbial consortia: process-based modeling of ANME-2 archaea and sulfate reducing bacteria interactions. *Biogeosciences* **5,** 1587–1599 (2008).

13. Alperin, M. J. & Hoehler, T. M. Anaerobic methane oxidation by archaea/sulfate-reducing bacteria aggregates: 1. thermodynamic and physical constraints. *Am. J. Sci.* **309,** 869–957 (2009).

14. Orphan, V. J., House, C. H., Hinrichs, K.-U., McKeegan, K. D. & DeLong, E. F. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl Acad. Sci. USA* **99,** 7663–7668 (2002).

15. Knittel, K., Lösekann, T., Boetius, A., Kort, R. & Amann, R. Diversity and distribution of methanotrophic archaea at cold seeps. *Appl. Environ. Microbiol.* **71,** 467–479 (2005).

16. Orphan, V. J., Turk, K. A., Green, A. M. & House, C. H. Patterns of ^{15}N assimilation and growth of methanotrophic ANME-2 archaea and sulfate-reducing bacteria within structured syntrophic consortia revealed by FISH-SIMS. *Environ. Microbiol.* **11,** 1777–1791 (2009).

17. Nauhaus, K., Boetius, A., Krüger, M. & Widdel, F. *In vitro* demonstration of anaerobic oxidation of methane coupled to sulphate reduction in sediment from a marine gas hydrate area. *Environ. Microbiol.* **4,** 296–305 (2002).

18. Thauer, R. K. Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO_2. *Curr. Opin. Microbiol.* **14,** 292–299 (2011).

19. Moran, J. J. *et al.* Methyl sulfides as intermediates in the anaerobic oxidation of methane. *Environ. Microbiol.* **10,** 162–173 (2008).

20. Milucka, J. *et al.* Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491,** 541–546 (2012).

21. Dolfing, J. The energetic consequences of hydrogen gradients in methanogenic ecosystems. *FEMS Microbiol. Ecol.* **101,** 183–187 (1992).

22. Schink, P. B. & Stams, A. J. M. in *The Prokaryotes* (eds Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 471–493 (Springer, 2013).

23. Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. Field and laboratory studies of methane oxidation in an anoxic marine sediment — evidence for a methanogen-sulfate reducer consortium. *Glob. Biogeochem. Cycles* **8,** 451–463 (1994).

24. Krüger, M., Wolters, H., Gehre, M., Joye, S. B. & Richnow, H.-H. Tracing the slow growth of anaerobic methane-oxidizing communities by (15)N-labelling techniques. *FEMS Microbiol. Ecol.* **63,** 401–411 (2008).

25. Schreiber, L., Holler, T., Knittel, K., Meyerdierks, A. & Amann, R. Identification of the dominant sulfate-reducing bacterial partner of anaerobic methanotrophs of the ANME-2 clade. *Environ. Microbiol.* **12,** 2327–2340 (2010).

26. Lovley, D. R. Electromicrobiology. *Annu. Rev. Microbiol.* **66,** 391–409 (2012).

27. Michelusi, N., Pirbadian, S., El-Naggar, M. Y. & Mitra, U. A stochastic model for electron transfer in bacterial cables. *IEEE J. Sel. Areas Comm.* **32,** 2402–2416 (2014).

28. Meysman, F. J. R., Risgaard-Petersen, N., Malkin, S. Y. & Nielsen, L. P. The geochemical fingerprint of microbial long-distance electron transport in the seafloor. *Geochim. Cosmochim. Acta* **152,** 122–142 (2015).

29. Meyerdierks, A. *et al.* Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ. Microbiol.* **12,** 422–439 (2010).

30. Wang, F.-P. *et al.* Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *ISME J.* **8,** 1069–1078 (2014).

31. Haroon, M. F. *et al.* Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500,** 567–570 (2013).

32. Strycharz-Glaven, S. M., Snider, R. M., Guiseppi-Elie, A. & Tender, L. M. On the electrical conductivity of microbial nanowires and biofilms. *Energy Environ. Sci.* **4,** 4366–4379 (2011).

33. Richardson, D. J. *et al.* The 'porin-cytochrome' model for microbe-to-mineral electron transfer. *Mol. Microbiol.* **85,** 201–212 (2012).

34. Okamoto, A., Hashimoto, K. & Nakamura, R. Long-range electron conduction of *Shewanella* biofilms mediated by outer membrane C-type cytochromes. *Bioelectrochemistry* **85,** 61–65 (2012).

35. Graham, R. C. & Karnovsky, M. J. The early stages of absorption of injected horseradish peroxidase in the proximal tubules of mouse kidney: ultrastructural cytochemistry by a new technique. *J. Histochem. Cytochem.* **14,** 291–302 (1966).

36. Litwin, J. A. Transition metal-catalysed oxidation of 3,3′-diaminobenzidine [DAB] in a model system. *Acta Histochem.* **71,** 111–117 (1982).

37. Welte, C. & Deppenmeier, U. Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. *Biochim. Biophys. Acta* **1837,** 1130–1147 (2014).

38. Summers, Z. M. *et al.* Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. *Science* **330,** 1413–1415 (2010).

39. Rotaru, A.-E. *et al.* Direct interspecies electron transfer between *Geobacter metallireducens* and *Methanosarcina barkeri*. *Appl. Environ. Microbiol.* **80,** 4599–4605 (2014).

40. Beal, E. J., House, C. H. & Orphan, V. J. Manganese- and iron-dependent marine methane oxidation. *Science* **325,** 184–187 (2009).

41. Kletzin, A. *et al.* Cytochromes c in Archaea: distribution, maturation, cell architecture and the special case of *Ignicoccus hospitalis*. *Front. Microbiol.* **6,** 439 (2015).

Extended Data Figure 1 | Image processing workflow for single cell correlation between FISH and nanoSIMS data sets. Representative example of data processing for an AOM consortium. **a**, Fiducial markers added to the FISH image. Marker points are shown in yellow, bacterial cells in red, archaeal cells in green. **b**, Corresponding fiducial markers identified on the nanoSIMS image. **c**, Overlay of the warped FISH image onto the nanoSIMS image, the transform function was defined by the points shown in **a** and **b**.

d, Overlay of the original FISH image (yellow) and the warped FISH image (blue) highlighting a slight offset which becomes significant at single-cell resolutions. **e**, Centroids of the hand-drawn ROIs displayed on the nanoSIMS image, bacteria in red, archaea in green. **f**, Inverse transform applied to the ROIs drawn on the nanoSIMS image, bringing the centroid coordinates into 'FISH space' where we have more accurate measurement of distances between points.
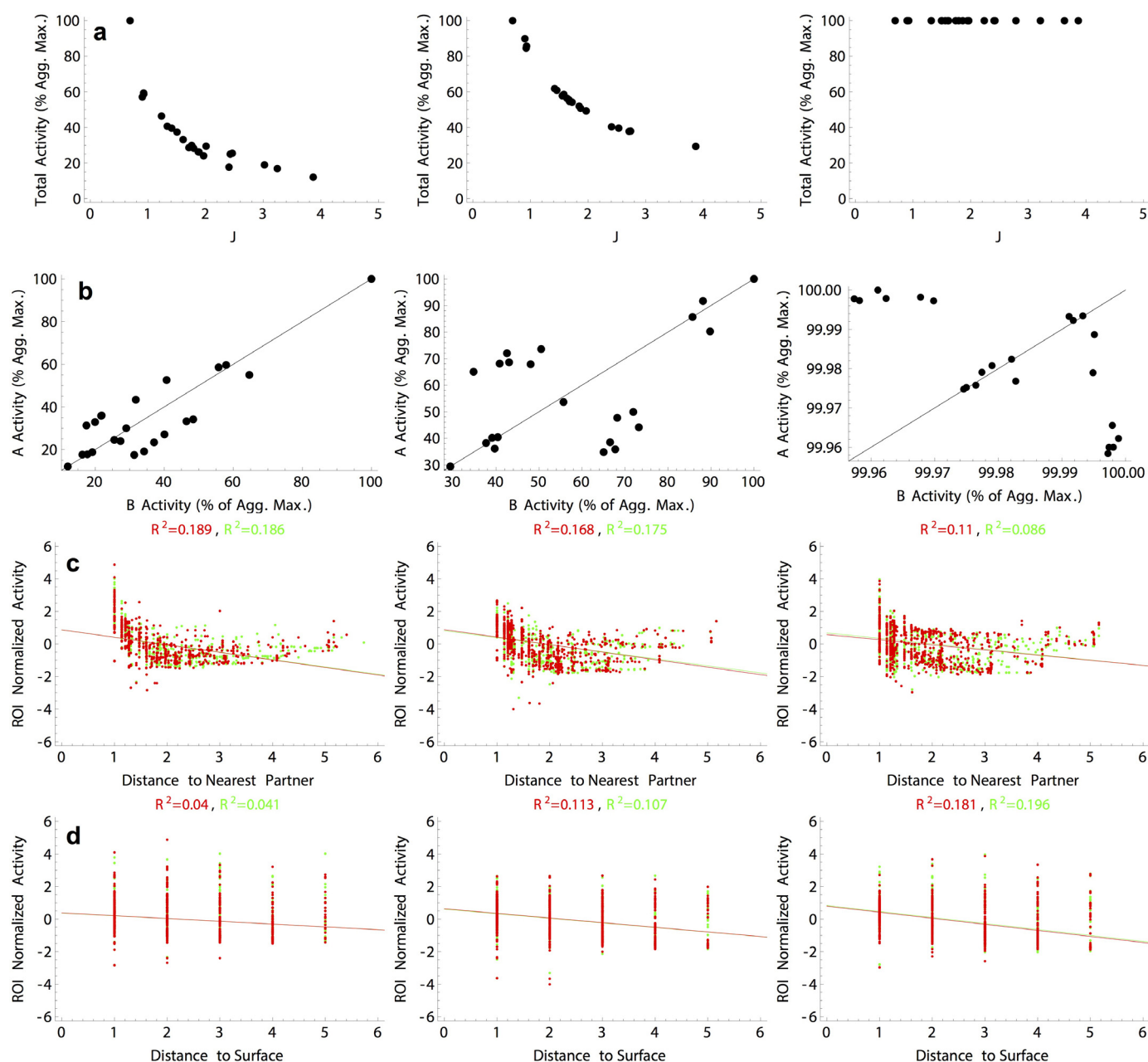
Extended Data Figure 2 | Spatial and geometric relationships for modelled aggregate geometries (well mixed to segregated) as a function of relative diffusivity (the ratio of growth rates to growth yields and diffusivity; see Supplementary Information) within the intermediate exchange model.
Slow diffusion is on the left (equivalent to roughly half the relative diffusivity of hydrogen compared to measured growth rates in our system) and fast on the right (equivalent to 10^3 times faster relative diffusivity than hydrogen compared with measured growth rates; see Supplementary Information). **a**, Total aggregate activity normalized to the group maximum as a function of the J spatial metric showing a strong dependency on geometry favouring well mixed (low J value) geometries under slow relative diffusion (left) and almost no relationship with J in fast-diffusion models (right). The average activity, normalized across all of the regimes rather than within a single regime, also changes dramatically from 0.002 to 0.99 as the relative diffusivity is increased. **b**, Total normalized archaeal population activity plotted against the total bacterial population activity within the same modelled aggregate. The total number of *in silico* consortia for rows **a** and **b** is 23. **c**, The normalized (*z*-score) activity for archaea (red) and bacteria (green) plotted against the distance to the nearest three partners. **d**, The *z*-score activity for archaea (green) and bacteria (red) plotted against the distance to environment-aggregate interface (that is, aggregate surface). In plots **c** and **d** the *r*-squared values for each correlation are given at the top of each plot in colours that correspond to the two cell types. The number of modelled *in silico* bacterial and archaeal cells from **c** and **d** plotted in the columns from left to right are: 1,138 bacterial and 1,162 archaeal cells; 1,163 bacterial and 1,137 archaeal cells; and 1,153 bacterial and 1,147 archaeal cells. As diffusion is increased in these models from left to right, the organisms within consortia become less dependent on each other and instead become less syntrophically coupled, relying on environmental exchange. This leads to the highest average activity rates per consortia (compare the top panel **a** to **b**).
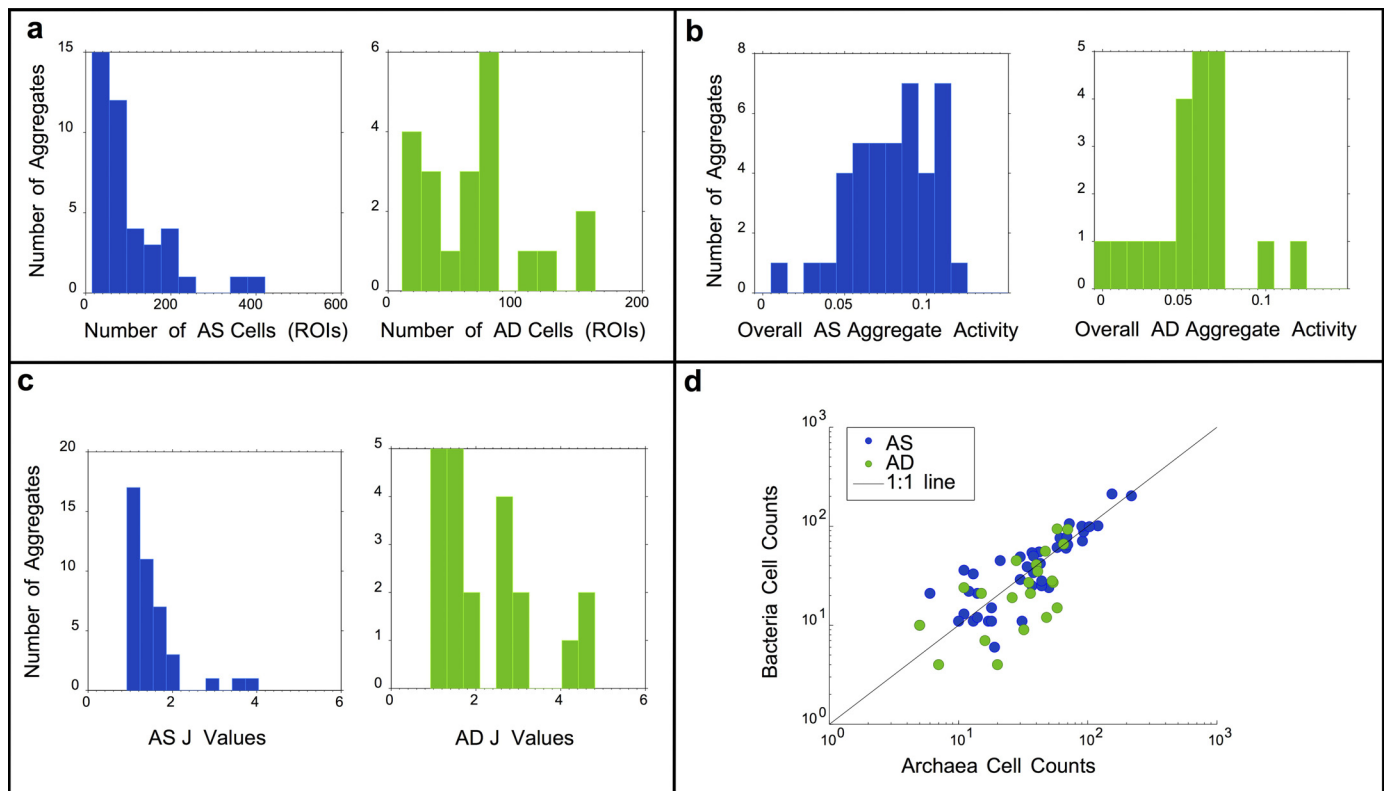
**Extended Data Figure 3 | Summary of aggregate characteristics.**
a, Histograms displaying the distribution of cell counts per aggregate for AS and AD consortia, blue and green respectively. **b**, Histograms displaying the average activity values for the AS and AD consortia, where anabolic activity is measured as fractional abundance of ^{15}N per cell. **c**, Histograms of the number of AS and AD consortia associated with different levels of spatial mixing between syntrophic partners represented by the spatial mixing metric 'J' (see Supplementary Discussion for details on this metric). **d**, One-to-one relation between bacterial and archaeal cell counts in the AS and AD consortia analysed in this study. For all panels, the data set consists of 41 AS and 21 AD consortia. The number of cells in each aggregate can be found in the Source Data.
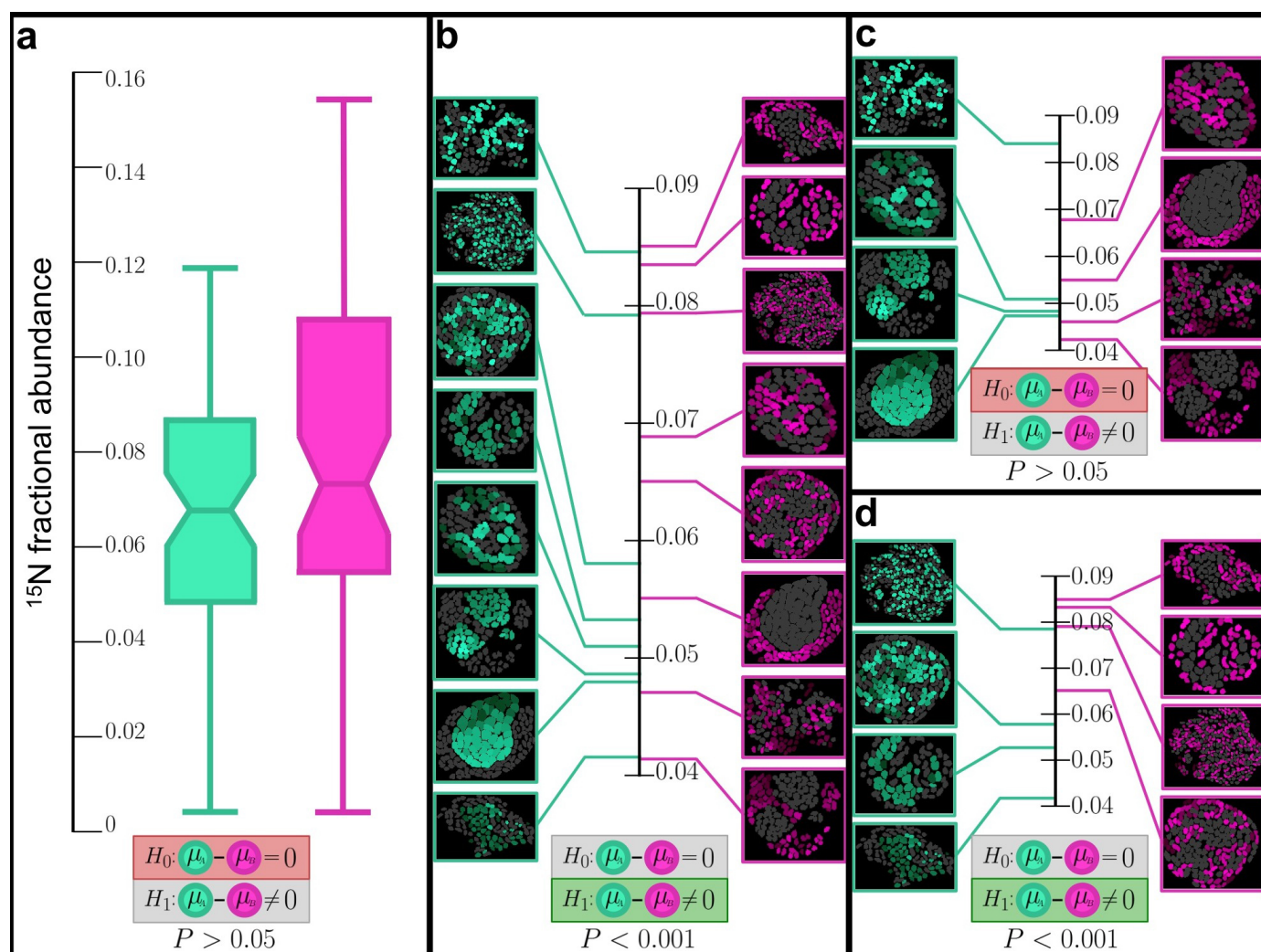
Extended Data Figure 4 | Illustration of the value of single-cell resolution activity analysis. a, Box plots showing the full range of archaeal and bacterial single-cell activities determined by $^{15}NH_4^+$ assimilation. The difference between the archaeal and bacterial mean activities across all aggregates ($n = 62$) is not significant (two sample t-test, $P > 0.05$). **b**, With our ability to quantify the activity for individual phylogenetically identified cells in AOM consortia, the average activity of the bacterial and archaeal populations within each consortium was revealed. Assessed at the level of paired populations, a significant difference in activity between the population of archaea and Deltaproteobacteria within aggregates is evident ($n = 62$, paired-sample t-test, $P < 0.001$). **c, d**, Adding phylogenetic resolution to this analysis by sub-grouping consortia based on their different deltaproteobacterial partners (AD and AS) reveals the difference between bacteria and archaea is only significant in the AS consortia ($n = 41$, paired-sample t-test, $P < 0.001$), while this population level offset in activity was not statistically supported within the AD group ($n = 21$, paired-sample t-test, $P > 0.05$), illustrating differential patterns in activity related to species membership. All axes represent ^{15}N fractional abundance. The 8 consortia images shown in panels **b–d** represent a subset of the total 62 consortia included in the analysis, with each image coloured by either archaeal ^{15}N enrichment on the left (green) or bacterial ^{15}N enrichment on the right (pink). The degree of brightness for each cell in the image reflects increasing levels of relative cellular ^{15}N enrichment and the average population value for ^{15}N fractional abundance is provided on the central axis.
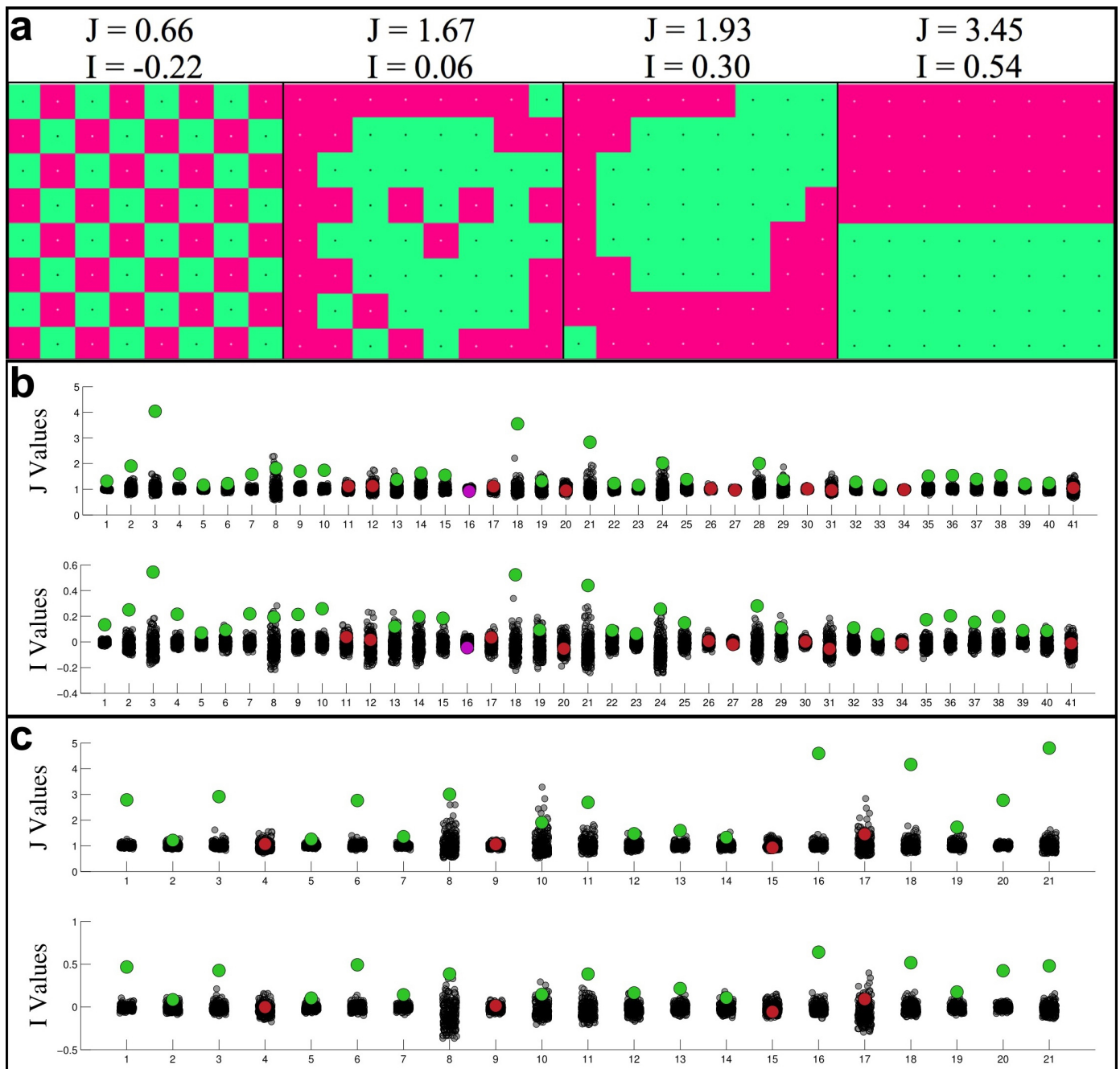
Extended Data Figure 5 | Evaluation of metrics for partner mixing. The degree of partner intermixing within an aggregate was calculated using two metrics (see Supplementary Information for detailed description of metrics). For the modified join metric (J), 1 represents random mixing, while for Moran's *I*, 0 represents random mixing. For both metrics increasing positive values represent increasing partner segregation and increasing negative values represent increasing ordered mixing (like a checkerboard). **a**, Examples of mock aggregates which were used to verify the behaviour of the two metrics. **b, c**, The determined values for either J or Moran's *I* are represented by the large coloured data points for each of the 41 AS aggregates or 21 AD aggregates analysed in this study, respectively. The black data points represent the values for J or Moran's *I* that were calculated in 300 permutation tests where the *x* and *y* coordinates of the archaea and bacteria cells were randomly reassigned. When the observed mixing was more segregated than 95% of the random permutation tests, the data points were coloured green and considered significant. Similarly, when the observed mixing was found to be more orderly mixed than 95% of the permutation tests the data points were coloured purple. When the observed mixing was found to be less extreme in either direction than 95% of the random test aggregates the data points were coloured red. The two metrics, while different in their formulation, gave very similar results. It is noteworthy that only a single aggregate contained cells that were more mixed than random. As expected, the permutation tests hover around the random mixing values for each metric, 1 and 0 for J and *I*, respectively.
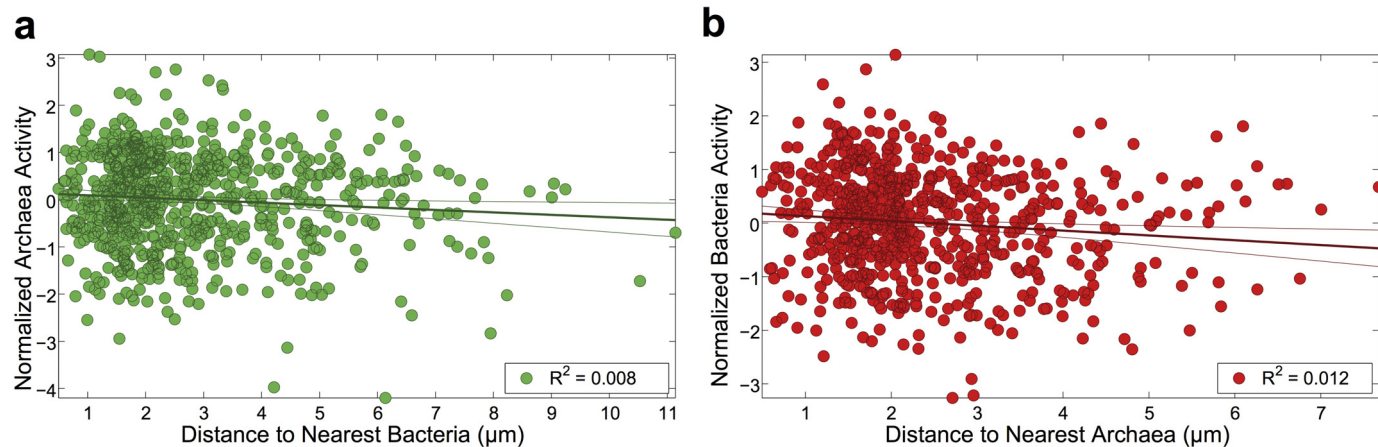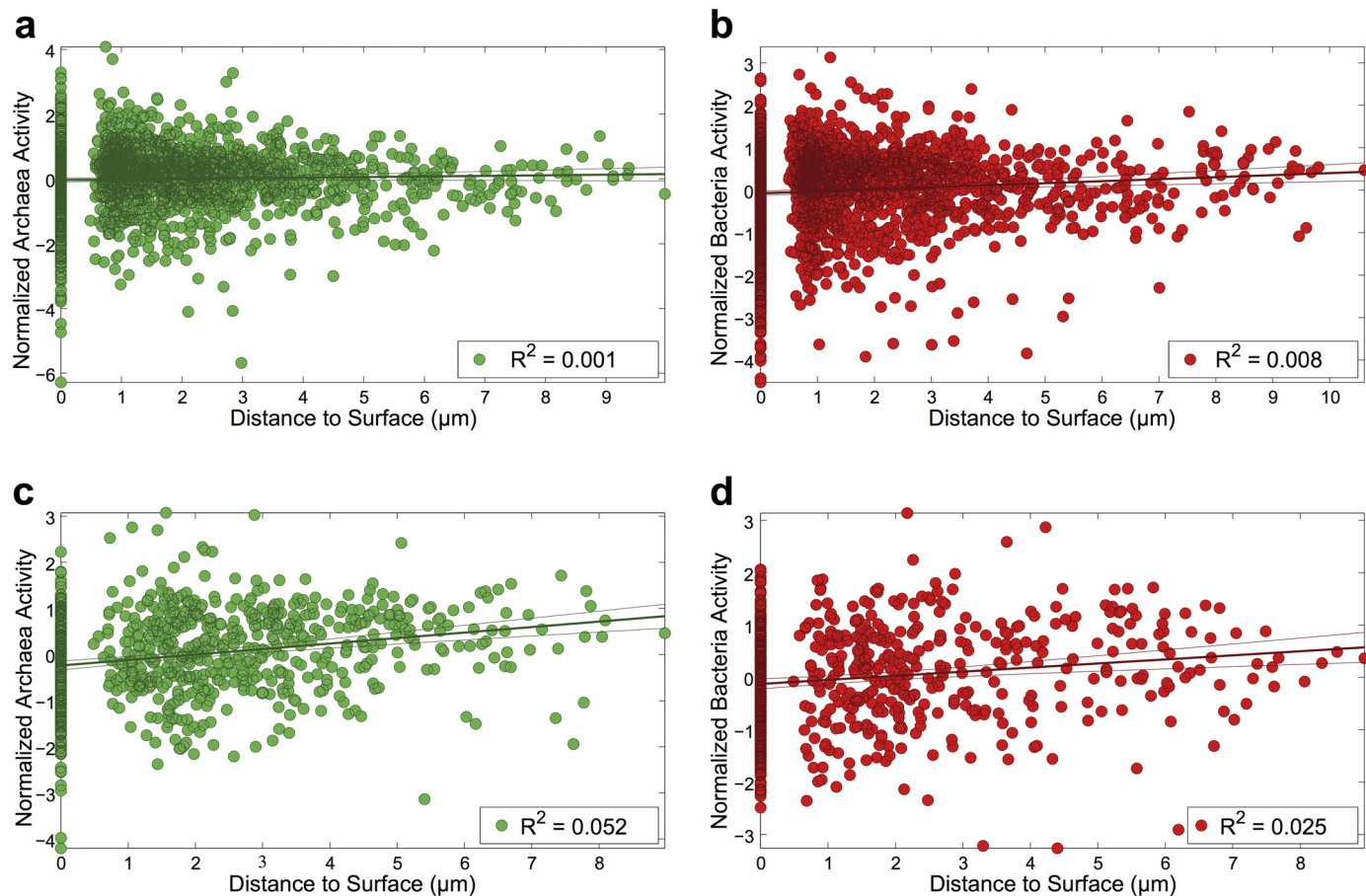
Extended Data Figure 6 | Insensitivity of cell activities to distance from nearest syntrophic partner for AD consortia. Plots displaying all ROIs analysed of a given type for consortia composed of ANME-2b or ANME-2c and Deltaproteobacteria. Normalized activity (Z-scores) were calculated within each aggregate to allow for comparisons between consortia with large differences in average cellular activity. **a**, Normalized activities of archaea ($n = 765$ cells) within AD consortia as a function of distance to nearest syntrophic partner. **b**, Normalized activities of bacteria ($n = 658$ cells) within AD consortia as a function of distance to nearest syntrophic partner. From this analysis, it appears that distance to nearest syntrophic partner does not account for a significant amount of the variation in cellular activity within a consortium. The R^2 values for linear regressions on the plotted data are shown in each panel. Dashed lines illustrate the 95% confidence intervals in slopes and intercepts of the linear regressions.
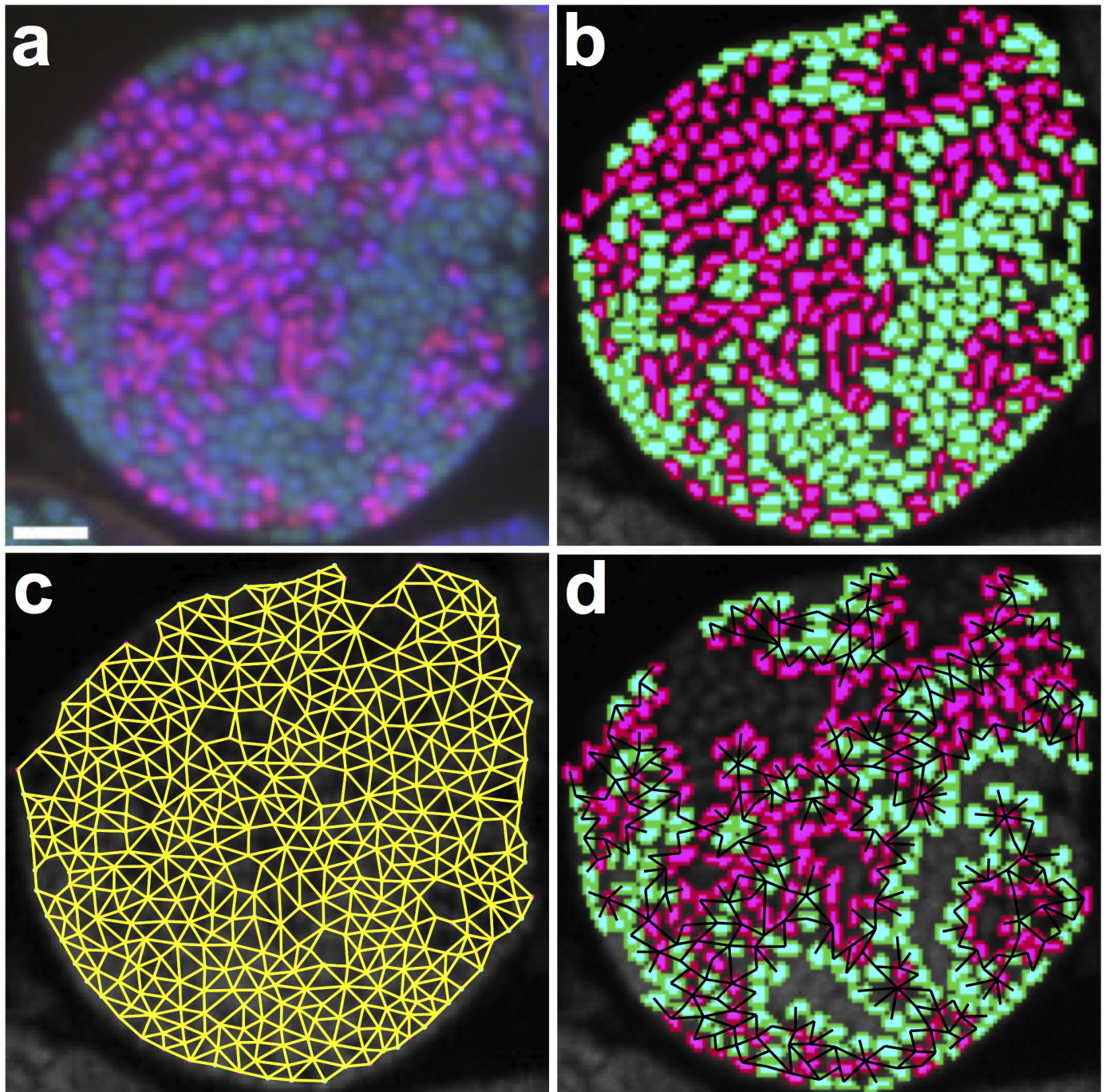
Extended Data Figure 7 | Schematic of network analysis for microbial consortia. a, FISH image of a representative ANME (green) and SRB (pink) consortium. **b**, Highlighted regions of interest false coloured by phylogenetic affiliation. **c**, Spheres of influence network of the consortia showing connectivity between cells. **d**, Identification of cells that share a border with a syntrophic partner (archaea adjacent to bacteria).

Extended Data Figure 8 | Insensitivity of cell activities to distance from surface. Plots displaying all ROIs analysed for a given population. Normalized activities (Z-scores) were taken within each consortium to allow for comparisons between aggregates with large differences in average cellular activity. **a**, Normalized activities of archaea within AS aggregates ($n = 1,967$ cells) as a function of distance to aggregate surface (that is, the external environment). **b**, Normalized activities of bacteria ($n = 2,063$ cells) within AS aggregates as a function of distance to aggregate surface. **c**, Normalized activities of archaea ($n = 765$ cells) within AD aggregates as a function of distance to aggregate surface. **d**, Normalized activities of bacteria ($n = 658$ cells) within AD aggregates as a function of distance to aggregate surface. From this analysis, the distance to the surface of the aggregate does not appear to explain a significant amount of the variation in cellular activity within each consortium. The R^2 values for linear regressions on the plotted data are shown in each panel. Dashed lines illustrate the 95% confidence intervals in slopes and intercepts of the linear regressions.
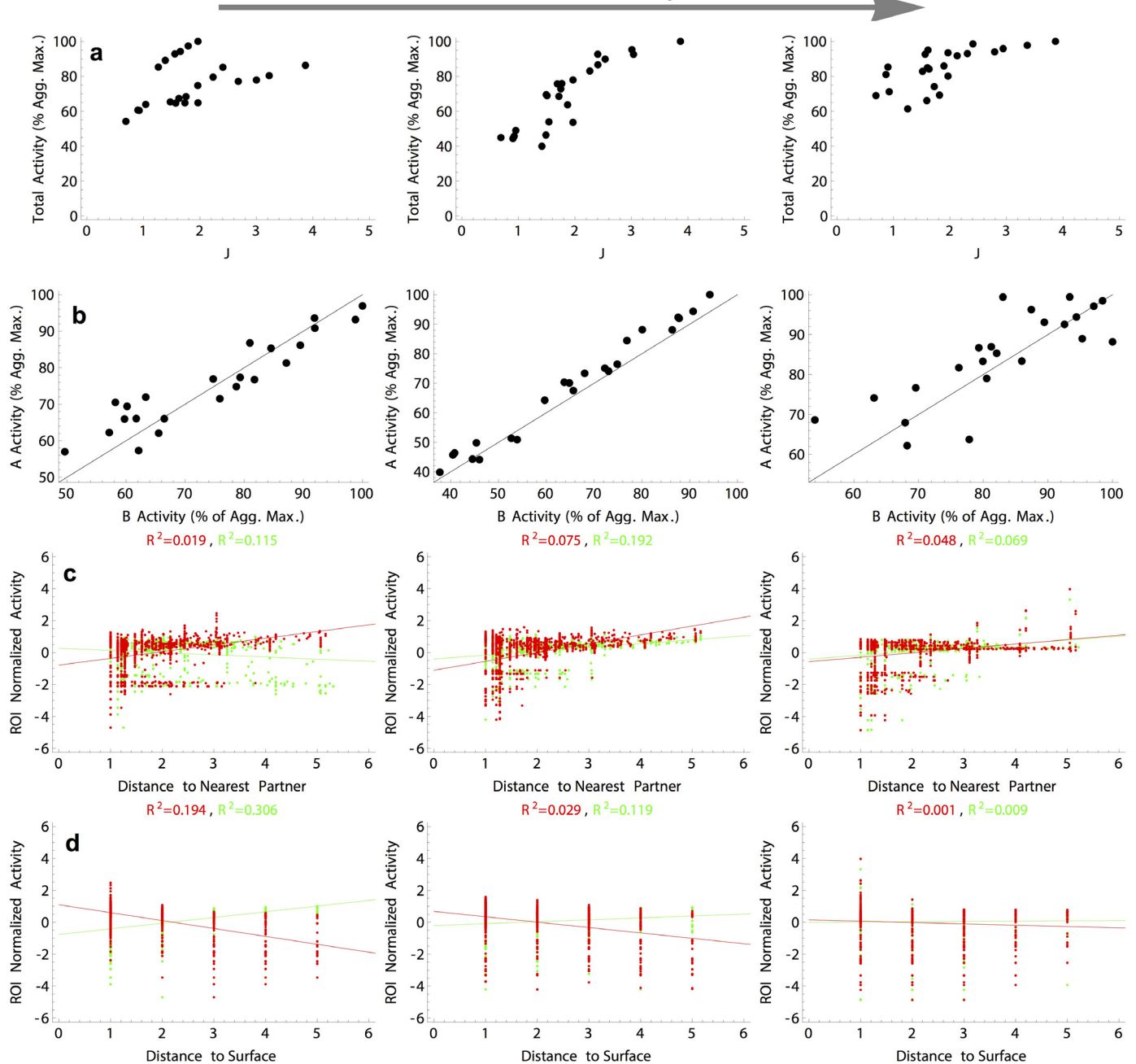
Extended Data Figure 9 | Spatial and geometric relationships for all modelled aggregate geometries as a function of relative conductivity within the direct electron transfer model. a, Total aggregate activity normalized to the group maximum as a function of the J spatial metric, from well-mixed (low J) to segregated (high J) aggregate geometries (23 *in silico* aggregates in total). These plots illustrate how the total activity of all of aggregate geometries changes with the relative conductivity, with less dependency on geometry observed at the fastest conductance rates. Compare to Extended Data Fig. 2: in the case of electron exchange presented here, the least mixed aggregates (high J) have the highest activity. This is because our conductive treatment of the aggregate relies on the global electric potential of each consortia, which is the strongest when the cells are spatially organized. **b**, Normalized archaeal activity plotted against the normalized bacterial activity within the same modelled aggregate. **c**, The normalized (*z*-score) activity for archaea (green) and bacteria (red) plotted against the distance to the nearest three partners. **d**, The *z*-score activity for archaea (green) and bacteria (red) plotted against the distance to environment-aggregate interface (aggregate surface). In plots **c** and **d** the *r*-squared values for each correlation are given at the top of each plot in colours that correspond to the two cell types. The number of modelled *in silico* bacterial and archaeal cells from **c** and **d** plotted in the columns from left to right are: 1,138 bacterial and 1,162 archaeal cells; 1,161 bacterial cells and 1,139 archaeal cells; and 1,134 bacterial cells and 1,166 archaeal cells.

ARTICLE

Structural basis for gene regulation by a B$_{12}$-dependent photoreceptor

Marco Jost[1], Jésus Fernández-Zapata[4], María Carmen Polanco[5], Juan Manuel Ortiz-Guerrero[5]†, Percival Yang-Ting Chen[1], Gyunghoon Kang[1], S. Padmanabhan[4], Montserrat Elías-Arnanz[5] & Catherine L. Drennan[1,2,3]

Photoreceptor proteins enable organisms to sense and respond to light. The newly discovered CarH-type photoreceptors use a vitamin B$_{12}$ derivative, adenosylcobalamin, as the light-sensing chromophore to mediate light-dependent gene regulation. Here we present crystal structures of *Thermus thermophilus* CarH in all three relevant states: in the dark, both free and bound to operator DNA, and after light exposure. These structures provide visualizations of how adenosylcobalamin mediates CarH tetramer formation in the dark, how this tetramer binds to the promoter −35 element to repress transcription, and how light exposure leads to a large-scale conformational change that activates transcription. In addition to the remarkable functional repurposing of adenosylcobalamin from an enzyme cofactor to a light sensor, we find that nature also repurposed two independent protein modules in assembling CarH. These results expand the biological role of vitamin B$_{12}$ and provide fundamental insight into a new mode of light-dependent gene regulation.

Light allows for photosynthesis and other essential light-dependent chemical reactions. Light also triggers photo-oxidative stress via generation of reactive oxygen species, which rapidly damage the cell[1]. Organisms in all domains of life produce proteins capable of sensing light. These biological photoreceptors participate in vision, harness light energy, regulate circadian clocks, and mediate gene expression and major developmental processes[2–5]. Different classes of photoreceptor proteins with a variety of chromophore cofactors exist to sense light over the visible and ultraviolet spectrum[6,7]. This group of photoreceptors was recently expanded by a new class that is widespread in bacteria and uses the vitamin B$_{12}$ derivative adenosylcobalamin (AdoCbl) as the chromophore to orchestrate light perception and response[8,9]. The prototypes of this class, the CarH-type photoreceptors, have been characterized in *Myxococcus xanthus*[8,9] and *T. thermophilus*[10] and typically regulate light-induced expression of carotenoid biosynthetic genes, which results in carotenoid-mediated protection against photo-oxidative damage[10–12]. This class of photoreceptors thereby allows bacteria to mitigate the detrimental effects of sunlight, a critical determinant of survival in light-exposed conditions, while avoiding unnecessary production of carotenoids in the absence of light.

The CarH-type photoreceptors consist of an amino (N)-terminal DNA-binding domain and a carboxy (C)-terminal AdoCbl-binding and oligomerization domain to directly sense light and regulate gene expression. In the dark, AdoCbl-bound CarH, a tetramer, binds to the promoter region of target genes to repress transcription. Exposure to blue, green, or ultraviolet light disrupts the photosensitive Co–C bond in AdoCbl, leading to tetramer disassembly, loss of operator-binding, and activation of gene expression (Fig. 1)[8]. Thus, AdoCbl, which is typically used as a cofactor for radical-based enzyme reactions[13,14], is now being used for a new biological function as a light sensor. Here, we sought to determine the structural basis for the functional repurposing of one of nature's most complex metallocofactors and for this new mode of light-dependent gene regulation.

CarH has a modular architecture

To visualize the CarH domain structure and the architecture of the repressor 'dark' state, we first determined two independent structures of AdoCbl-bound CarH from *T. thermophilus* to 2.15 Å and 2.80 Å resolution (Extended Data Table 1). To prevent cleavage of the light-sensitive AdoCbl Co–C bond, we performed all crystallization experiments under red light and all diffraction data collection at $T = 100$ K. Both single-crystal ultraviolet–visible (UV–vis) spectra and the electron density confirmed that AdoCbl remained intact with a Co–C bond length of 2.0 Å (Extended Data Fig. 1), indicating that the structures are in the 'dark' state. Each monomer of CarH has a modular three-domain architecture with an N-terminal winged-helix DNA-binding domain followed by the light-sensing domain, composed of a four-helix bundle and a C-terminal Rossmann-fold cobalamin (Cbl)-binding domain (Fig. 2a).

The DNA-binding domain is structurally similar to those of MerR family transcription factors[15] (Fig. 2b) and to that of the AdoCbl-independent CarH paralogue CarA[8,16], featuring a canonical recognition helix and a β-hairpin wing for DNA binding (Fig. 2a). In our DNA-free structures, the DNA-binding domain samples different orientations for the different CarH protomers in the asymmetric unit, enabled by a flexible linker region and stabilized by crystal lattice contacts (Extended Data Fig. 2).

In contrast to the flexible DNA-binding domains, the four-helix bundle and the Rossmann domain are structurally rigid, together forming a module that binds the AdoCbl light sensor. AdoCbl is sandwiched between the four-helix bundle, which interacts with the upper axial 5′-deoxyadenosyl (5′-dAdo) ligand, and the Rossmann-fold domain, which binds the Cbl lower face in the base-off/His-on mode with the side chain of His177 displacing the Cbl dimethylbenzimidazole base (Fig. 2a, d and Extended Data Fig. 1b). Instead of closely resembling an AdoCbl-dependent enzyme, the CarH light-sensing domain is structurally homologous to the methylcobalamin
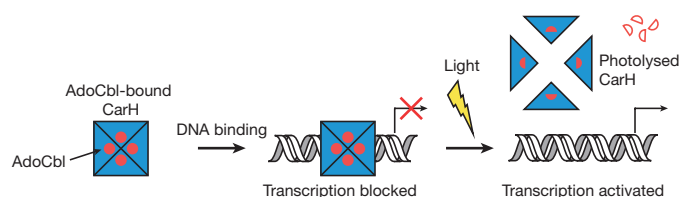
Figure 1 | Schematic of CarH-mediated light-dependent gene regulation. Structures of all three relevant states are reported herein. Red circles depict AdoCbl, filled red semicircles photolysed Cbl, and open red semicircles 4′,5′- anhydroadenosine, the recently identified photolysis product of CarH-bound AdoCbl[33]. See main text for details.

(MeCbl)-binding module of methionine synthase MetH (Fig. 2b)[17], even though the AdoCbl 5′-dAdo group is much bulkier than the MeCbl methyl group in MetH. Modelling AdoCbl into MetH leads to several steric clashes (Fig. 2c), but in CarH, a small but important shift of 2.5 Å of the four-helix bundle enlarges the cavity on the Cbl upper face (Fig. 2d, e). Additionally, four hydrophobic residues at the Cbl upper face in MetH are replaced in CarH, providing the 5′-dAdo group a larger binding pocket (Leu715→Val138), a hydrogen bonding interaction (Val718→Glu141), and more polar environment in general (Val719→His142, Phe708→Trp131) (Fig. 2d, f). Although Trp is larger than Phe, the orientation of the Trp side chain on the side of the upper ligand rather than directly above perfectly accommodates the 5′-dAdo group (Fig. 2c, d). Notably, Trp131, Glu141, and His142 are highly conserved in CarH homologues, suggesting that this mode of AdoCbl binding is conserved as well (Extended Data Fig. 3).

CarH is a dimer–of–dimers type tetramer

AdoCbl-bound CarH is a tetramer in the crystal structure (Fig. 3), consistent with results from size-exclusion chromatography (SEC) and analytical ultracentrifugation[8,18]. Four light-sensing domains comprise the core of the tetramer (Fig. 3a–d) with the DNA-binding domains extending outward (Fig. 3e, f). The core has a dimer-of-dimers architecture, with each constituent dimer composed of two CarH light-sensing domains in a head-to-tail orientation (Fig. 3a). The extensive head-to-tail dimer interface is formed by the four-helix bundles and the Cbl-binding domains and features a solvent-buried area of 1,430 Å^2 on each protomer as well as numerous hydrogen bonds and ionic interactions involving various side chains and the 5′-dAdo group of AdoCbl (Fig. 3a, b). Two such head-to-tail dimers assemble to a tetramer in a staggered fashion (Fig. 3c–f). This dimer–dimer interface is again formed by the light-sensing domains, creating a buried surface area of 1,590 Å^2 on each of the two head-to-tail dimers, whereas the four DNA-binding domains are positioned on

the surface of the tetramer and only make minor contributions to the interface (Fig. 3e, f). It was previously demonstrated by SEC that CarH lacking the DNA-binding domain still forms tetramers[8], and here we find that CarH adopts the same tetramer architecture when the DNA-binding domains are proteolytically removed during crystallization (Fig. 3c, e). Thus, the light-sensing domains appear to mediate tetramerization, positioning the DNA-binding domains on the surface to engage DNA.

To confirm the CarH tetramer architecture and the mode of AdoCbl binding, we mutated residues in the AdoCbl binding site, at the head-to-tail dimer interface, and at the dimer–dimer interface and analysed these mutants using SEC and electrophoretic mobility shift assays (EMSAs) (Extended Data Figs 3 and 4). Non-conservative mutations in the binding site for the 5′-dAdo group (W131A, E141A; Fig. 2d) impaired tetramer formation and mutations near the head-to-tail dimer interface (H142A, D201R; Fig. 3b) completely abolished it. Remarkably, the most drastically adverse H142A and D201R mutations also appeared to impair AdoCbl binding (Extended Data Fig. 4a, c; absorbance traces at 522 nm). Moreover, DNA binding affinity weakened with decreased ability to form tetramers in the W131A, E141A, H142A, and D201R mutants (Extended Data Figs 3a and 4g). For comparison, we also introduced the conservative W131F mutation, which behaved like wild-type (WT) CarH in its oligomerization and DNA binding properties (Extended Data Fig. 4a, g). The inability of the D201R mutant protein to oligomerize is consistent with the observed Asp201/Arg176 interaction playing an important role in stabilizing the head-to-tail dimer interface (Fig. 3a). We expected a second compensatory mutation, R176D or R176E, to restore the interaction and, indeed, the R176D/D201R and R176E/D201R double mutants could form tetramers (Extended Data Fig. 4c) and bind to DNA, albeit less efficiently than WT CarH (Extended Data Fig. 4g). Finally, replacing Gly160 and Gly192 at the CarH dimer–dimer interface (Fig. 3d) by the bulkier Gln resulted in dimers in the presence of AdoCbl, a form previously never observed for WT CarH (Extended Data Fig. 4e). Although both these mutants bound to the DNA probe in the dark, they formed a smaller size (higher mobility) complex than WT CarH, suggesting that their binding mode is distinct (Extended Data Fig. 4g). Both mutants furthermore bound DNA with reduced affinity and cooperativity compared with WT CarH (Extended Data Fig. 4h). Together, these results are consistent with the observed CarH tetramer architecture and indicate that this architecture is critical for DNA binding.

CarH binds the promoter −35 element

To reveal the mode of transcriptional repression, we next sought to visualize CarH bound to its cognate DNA operator. The CarH operator lies within a 110-base-pair (bp) segment of the intergenic region between *carH* and the carotenogenic *crtB*[8,10]. Using systematically truncated DNA probes in EMSAs (Extended Data Fig. 5a, b), this operator
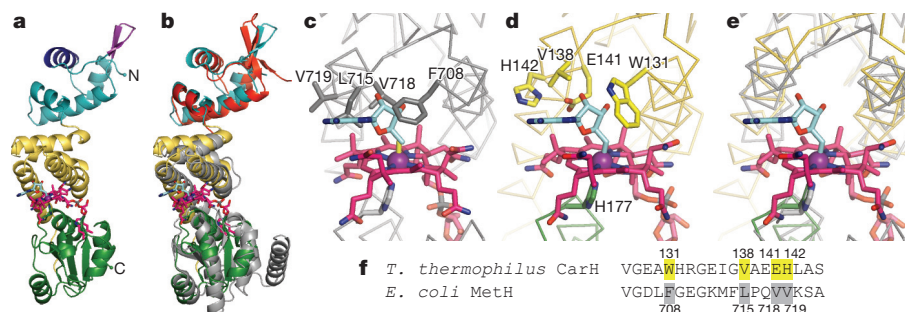

Figure 2 | Structure of CarH protomer and comparison with MetH. a, CarH protomer coloured by domain: N-terminal DNA-binding domain (cyan) with recognition helix (dark blue) and β-hairpin wing (purple) highlighted; central four-helix bundle (yellow); C-terminal Cbl-binding domain (green). AdoCbl shown with Cbl carbons in pink, 5′-dAdo group carbons in cyan, cobalt in purple. **b,** Overlay of CarH protomer with Cbl-binding module of MetH (grey, PDB accession number 1BMT[17]) and BmrR DNA-binding domain (red, PDB accession number 1EXJ[38]). **c,** MetH binds MeCbl (methyl group in yellow); modelling 5′-dAdo (cyan) results in steric clashes. **d,** CarH accommodates AdoCbl through several substitutions compared with MetH. Cobalt-coordinating His in green. **e,** Superposition of MetH and CarH, highlighting shift of the four-helix bundle between MetH (grey) and CarH (yellow). **f,** Alignment of CarH and MetH sequences involved in binding the Cbl upper face, highlighting substitutions.
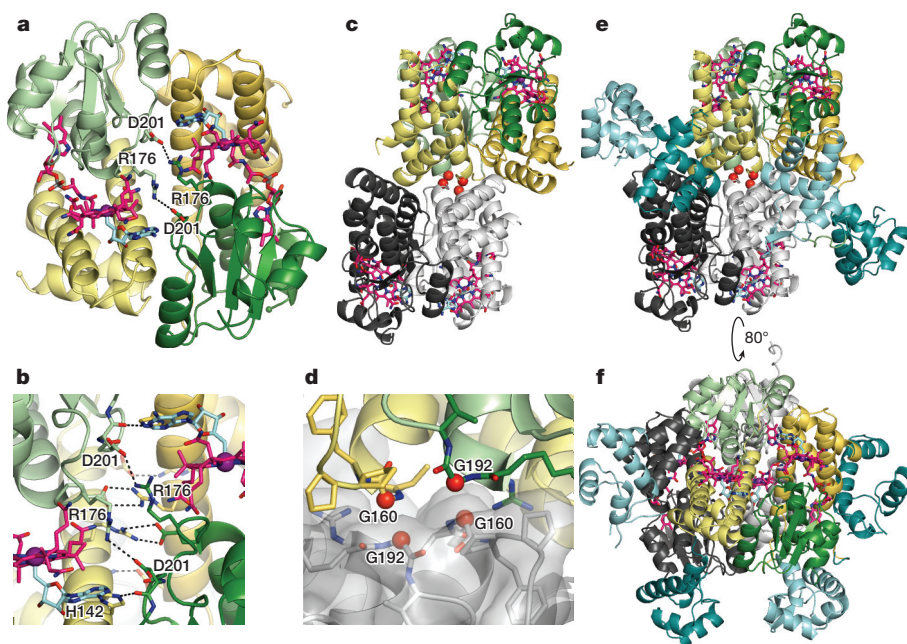
Figure 3 | CarH oligomerization. a, CarH protomers arranged in a head-to-tail dimer, coloured by domain (helix bundle: yellow; Cbl-binding domain: green) with left protomer shown in lighter colours. AdoCbl shown as in Fig. 2a. DNA-binding domains hidden for clarity. **b,** Close-up of selected residues at the dimer interface. **c,** Core of CarH tetramer, assembled from two head-to-tail dimers. Top dimer is coloured as in **a**; bottom dimer is coloured in black and grey. Gly160 and Gly192 at the dimer–dimer interface are shown as red spheres. Structure is from a sample of CarH that degraded during crystallization and lacks the DNA-binding domains (crystal form 2, see Methods). **d,** Close-up of Gly160 and Gly192 (red spheres) from two protomers at the dimer–dimer interface. **e,** Tetramer of full-length CarH including DNA-binding domains (crystal form 3), coloured as in **c**. DNA-binding domains of coloured dimer are shown in dark cyan, those of grey dimer are shown in light cyan. **f,** Additional view of CarH tetramer, revealing how DNA-binding domains are positioned on the protein surface.

sequence was narrowed down to a 30-bp segment. Diffraction quality crystals of AdoCbl-bound CarH were obtained in complex with a blunt-ended 26-bp DNA segment (−47 to −22 relative to *carH* transcription start site, Extended Data Fig. 5a), allowing us to determine the crystal structure of DNA-bound CarH to 3.89 Å resolution (Fig. 4, Extended Data Fig. 6a–h and Extended Data Table 1).

The structure revealed a unique mode of DNA binding involving three of the four DNA-binding domains of tetrameric CarH (Fig. 4a). The fourth DNA-binding domain is disordered and not visible in the electron density. The overall architecture of the tetramer is the same before and after DNA binding except for a reorientation of the DNA-binding domains (Extended Data Fig. 6e). All three DNA-binding domains face the same way on the DNA segment and bind to a set of three 11-bp repeats with a consensus sequence (A/G)A(G/C)(A/C)T(A/G/T)

(T/G)ACA(A/T) (Fig. 4a). This parallel orientation is stabilized by specific interactions between adjacent DNA-binding domains (Extended Data Fig. 6f). The central DNA-binding domain comes from one head-to-tail dimer, whereas the two flanking DNA-binding domains come from the second head-to-tail dimer (Extended Data Fig. 6g, h). These structures suggest that the two individual head-to-tail dimers would bind to DNA less tightly, consistent with the reduced affinity and cooperativity for the dimeric G160Q and G192Q CarH mutants (Extended Data Figs 3a and 4g, h). CarH rendered monomeric by light exposure[8,18] or mutagenesis (H142A, D201R) binds DNA with even further reduced affinity (Extended Data Figs 3a and 4g).

To obtain support for this unusual DNA binding mode, hydroxyl radical and DNase I footprinting were used to examine the regions of DNA protected by CarH. The DNase I footprint, which was obtained
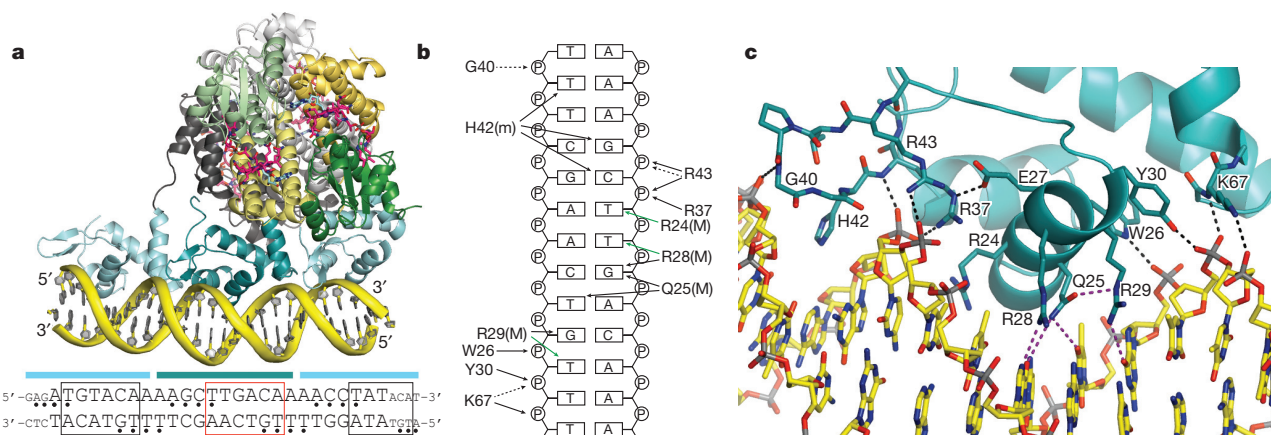

Figure 4 | CarH DNA binding. a, CarH tetramer bound to a 26-bp DNA segment (yellow). CarH is shown in ribbons with one head-to-tail dimer in green (Cbl-binding domains) and yellow (helix bundles) and the other dimer in grey. AdoCbl shown as in Fig. 2a. DNA-binding domains are shown in cyan. Sequence of DNA segment used for crystallization (larger font) as well as flanking sequences in the operator (smaller font) are shown below. Cyan bars indicate base pairs covered by each DNA-binding domain. Base pairs covered by the recognition helix are boxed; red box highlights the promoter −35 element. Nucleotides protected from hydroxyl radical cleavage are indicated by bullets. The orientation of the DNA in the structure was confirmed by heavy atom labelling (Extended Data Fig. 6b–d). **b,** Schematic of CarH–DNA

interactions for the central DNA-binding domain, denoted as follows: black arrows, hydrogen bonds/electrostatic interactions; green arrows, van der Waals interactions; solid lines, contacts from protein side chains; dashed lines, contacts from main chain; (M), contacts in the DNA major groove; (m), contacts in the DNA minor groove. **c,** Close-up of interactions between CarH (cyan) and DNA (yellow). Hydrogen bonds and ionic interactions to the phosphate backbone are shown as black dashed lines. Contacts to DNA bases are shown in purple. Side-chain orientation is not unambiguous owing to the modest resolution of this structure, but many of the contacts shown are supported by mutagenesis (Extended Data Fig. 7).

using a longer DNA segment (130 bp) than the one used in the crystal structure (26-bp), still matches the crystal structure footprint (Extended Data Fig. 5a, c). Additionally, three evenly spaced 4-bp hydroxyl radical footprints are observed on both the sense and the antisense strand (Extended Data Fig. 5c) that correlate with where the 'wings' of the three DNA-binding domains contact the minor groove (Fig. 4a). Taken together, the size of the DNase I footprint, the presence of three hydroxyl radical footprints, and the observation that the DNA sequence contains three repeats (see above), suggest that CarH binds to DNA using three of its four DNA-binding domains.

To determine whether all three repeats are important for high-affinity binding of CarH to DNA, we tested the effect of mutating DNA bases in the CarH operator (Extended Data Fig. 5d, e). Mutating dinucleotides in any single repeat only led to a small decrease in affinity as evidenced by the intense retarded bands for the CarH–DNA complex and the small amounts of free DNA in EMSAs (mutants 1–3 or 8–10 in Extended Data Fig. 5e). In contrast, simultaneously mutating dinucleotides in any two of the three repeats or in all three repeats almost completely abolished DNA binding (mutants 4–7 and 11–14 in Extended Data Fig. 5e). As a control, we also mutated DNA bases in the operator that CarH does not contact directly, and as expected, WT CarH bound to these mutants with similar affinity (mutants 15–18 in Extended Data Fig. 5e). Given that the results of mutations were similar for each of the three repeats, it appears that all three repeats are important in determining CarH–DNA affinity.

Each DNA-binding domain forms hydrogen bonds and electrostatic interactions to the phosphate backbone, contributed from both the peptide backbone and the side chains of Trp26, Tyr30, Arg37, Arg43, and Lys67 (Fig. 4b, c), and each domain also inserts His42 of its β-hairpin 'wing' into the DNA minor groove. Finally, each DNA-binding domain places its recognition helix in the DNA major groove, where it recognizes a 6-bp stretch (Fig. 4a) using specific hydrogen bonds from the side chains of Gln25, Arg28, and Arg29 (Fig. 4b, c). Strikingly, the major groove sequence occupied by the central DNA-binding domain spans the promoter −35 element (TTGACA, red box in Fig. 4a) for the major σ^A/σ^{70}-associated bacterial RNA polymerase. Thus, these structures reveal the mechanism of transcriptional repression: CarH occupies the −35 element, blocks access by the RNA polymerase-σ^A holoenzyme, and thereby prevents transcription initiation.

To verify the observed mode of DNA binding, we generated the Q25A, R29A, Y30A, H42A, and R43A CarH mutants and tested their DNA binding capacity using EMSAs. Mutating the conserved Arg29 or Arg43 to Ala abolished DNA binding but did not affect AdoCbl-dependent tetramerization (Extended Data Figs 3a and 7), consistent

with a role of these residues in DNA binding. Notably, the Q25A mutant retained DNA-binding capacity, indicating that Gln25 could be more important for mediating DNA specificity than affinity (Extended Data Fig. 7c). Finally, the H42A and Y30A mutants only showed mildly reduced affinity, suggesting that their interactions are not essential for DNA binding (Extended Data Figs 3a and 7c).

Light triggers helix bundle movement

Finally, to examine how light exposure causes tetramer disassembly, we determined the crystal structure of light-exposed CarH to 2.65 Å resolution (Fig. 5, Extended Data Fig. 8a, b and Extended Data Table 1). The structure contains monomeric CarH with bound Cbl but without the 5′-dAdo group, which dissociated as a consequence of light exposure (Fig. 5a). The four-helix bundle and the Cbl-binding domain individually do not exhibit major conformational changes compared with the dark AdoCbl-bound structure. However, the orientation of the helical bundle relative to the Cbl-binding domain has changed drastically with a >8 Å displacement (Fig. 5a, b). This helix bundle movement would disrupt the head-to-tail dimer interface (Fig. 5c), leading to tetramer disassembly, dissociation from DNA, and transcriptional activation.

Tetramer disassembly is triggered by loss of the AdoCbl 5′-dAdo group: its presence in the AdoCbl-bound structure blocks movement of the helix bundle, owing in large part to the positioning of W131 against the upper Cbl ligand (Fig. 2d), keeping the CarH protomers in the extended 'upright' conformation required for tetramerization. Loss of the 5′-dAdo group leaves a large cavity on the Cbl upper face (Fig. 5d), prompting movement of the helix bundle to occupy this void and cover the Cbl (Fig. 5e).

Strikingly, the helix bundle motion brings His132 from the protein surface to the Cbl upper face, where it binds to the cobalt to form bis-His ligated Cbl (Fig. 5d, e and Extended Data Fig. 8b). Such bis-His ligation, common for haems, has not been reported for Cbl, although bis-His Cbl ligation was recently proposed on the basis of mass spectrometry for the Cbl-dependent transcription factor AerR[19]. We therefore validated formation of bis-His-ligated Cbl using UV–vis spectroscopy. Spectra of light-exposed WT CarH and a H132A mutant, which is unable to form the bis-His ligation, resemble those of free Cbl with two or one nitrogen-based ligands[20], respectively (Extended Data Fig. 8c, d). In contrast, the spectra of the AdoCbl-bound proteins are identical (Extended Data Fig. 8e). These results provide unambiguous evidence for a bis-His-ligated Cbl in light-exposed CarH and suggest that this mode of coordination might be used more frequently in non-haem proteins.

Notably, both WT and H132A CarH undergo light-dependent tetramer disassembly, indicating that bis-His ligation is not required
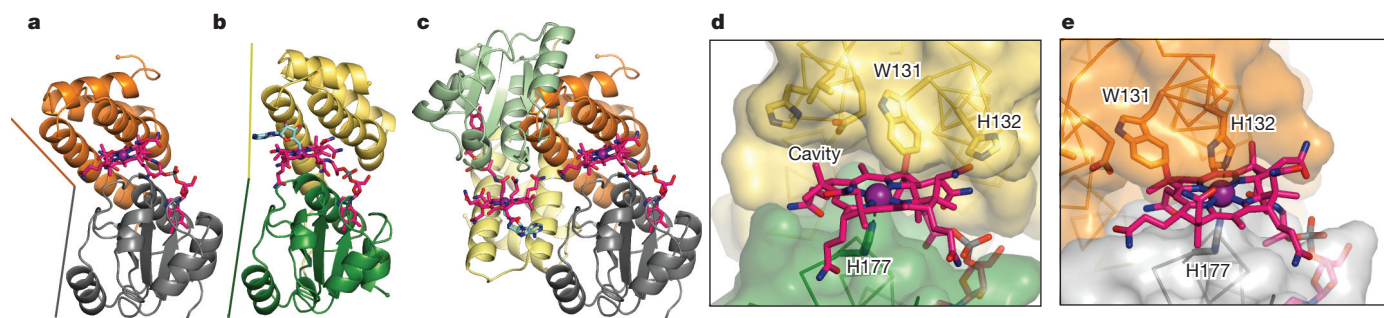
Figure 5 | Light-induced conformational changes in CarH. a, Structure of light-exposed CarH, with the helix bundle in orange and Cbl-binding domain in grey. DNA-binding domain is hidden for clarity (see Extended Data Fig. 8a). Cbl shown with carbons in pink and cobalt in purple. **b,** Structure of a CarH protomer in the dark state, shown with helix bundle in yellow, Cbl-binding domain in green, and 5′-dAdo group in cyan. Coloured lines highlight domain orientations. **c,** Light-induced helix bundle movement causes tetramer disassembly. Shown is a head-to-tail dimer of CarH in the dark state,

but the right protomer is replaced by the structure of light-exposed CarH to show the steric clash. **d,** Departure of the 5′-dAdo group after light exposure leaves a large cavity on the Cbl upper face. The helix bundle (yellow) and the Cbl-binding domain (green) are shown in surface representation with selected residues shown as sticks. **e,** Helix bundle movement fills the cavity at the Cbl upper face and brings His132 to the cobalt, where it occupies the open coordination site. Colouring as in **a.**

for disassembly (Extended Data Fig. 8f, g). However, Cbl dissociation after light exposure is faster for H132A CarH than for WT CarH, which forms a very tight and stable complex with the photolysed Cbl, as indicated by the relative abilities of the protein to be reconstituted with fresh AdoCbl (Extended Data Fig. 9a, b) and by the observation of a CarH:Cbl adduct in mass spectrometry for WT CarH but not for the H132A mutant (Extended Data Fig. 9c, d). Thus, the bis-His ligation could be important to retain the Cbl cofactor after photolysis.

Discussion

Impressively, CarH-type photoreceptors are found in hundreds of bacterial genomes, including bacteria that uptake rather than biosynthesize AdoCbl[8]. CarH is distinguished from most known classes of photoreceptors (with the exception of some light-oxygen-voltage (LOV)-type photoreceptors[21-23]) in that it can bind to DNA directly, instead of requiring additional proteins for gene regulation. Beyond gene regulation, the CarH light-sensing domains can be found fused to effector domains such as histidine kinases and in stand-alone modules that could undergo light-dependent protein–protein interactions[8]. This versatility probably explains the broad distribution of CarH-like proteins in bacteria.

The use of AdoCbl as a light-sensing chromophore by CarH is biologically unprecedented. AdoCbl is structurally and photochemically distinct from known photoreceptor chromophores such as bilin[24], flavins[25-27], retinal[4], or Trp side chains[28]: light exposure leads to breakage of a covalent Co–C bond, whereas other chromophores undergo less drastic changes such as light-induced electron transfer or cis–trans isomerizations. In all cases, however, light energy is ultimately harnessed to drive a large-scale conformational change, highlighting the convergence of different light sensing mechanisms. AdoCbl was previously best-known as a cofactor for radical-based enzyme reactions, in which reversible homolytic cleavage of the Co–C bond provides access to the 5′-dAdo radical for catalysis[13], and as a modulator of gene expression via riboswitches[29-32]. Our structures now allow us to visualize how AdoCbl is repurposed as a light sensor in CarH: in the dark, the AdoCbl 5′-dAdo group acts as a molecular doorstop that keeps CarH protomers in an extended 'upright' conformation for tetramerization, and light exposure triggers collapse into a kinked conformation. Whereas AdoCbl photolysis is an unwanted side reaction in enzyme catalysis because it leads to cofactor inactivation, in CarH this light sensitivity is harnessed to drive a light-dependent gene expression switch and a change in physiology. Remarkably, use of AdoCbl in this alternative function appears to come with a safeguard: the product of CarH-bound AdoCbl photolysis is not a 5′-dAdo radical, but rather 4′,5′- anhydroadenosine[33], which differs by one proton and one electron, and cannot cause radical damage. Thus, AdoCbl now joins the list of enzyme cofactors that have been repurposed as sensors; a list that already includes flavins (as light sensors in LOV, blue light sensor using flavin (BLUF), and cryptochrome photoreceptors[25-27]) and haems (as sensors of oxygen and other small molecules[34]).

Our CarH structures additionally reveal the functional repurposing of two different protein modules. The CarH light-sensing domain mirrors the Cbl-binding module of methionine synthase MetH, an enzyme that uses a MeCbl intermediate in the transfer of a methyl group from methyltetrahydrofolate to homocysteine, generating tetrahydrofolate and methionine[35]. But whereas MetH uses its helix bundle to position Phe708 over the MeCbl methyl group and protect it from photolysis[36], CarH, enabled by specific substitutions at the Cbl upper face, uses this fold as an AdoCbl-binding light-sensing domain, in which the corresponding Trp131 senses the presence of the 5′-dAdo group and transmits the signal of AdoCbl photolysis by leading a conformational change of the helix bundle. Furthermore, whereas the Cbl-binding module of MetH is embedded in a 136 kDa multi-domain protein and juxtaposed to different substrate-binding domains via transient

domain–domain interactions during a catalytic cycle[37], the light-sensing domain of CarH is used to assemble a tetramer that is stable enough to occlude the −35 element from RNA polymerase. Thus, this module has been repurposed from a methyl group carrier in primary metabolism to a light-sensing modulator of oligomerization state.

Similarly, the CarH DNA-binding domain resembles those of MerR-type transcription factors such as MerR, BmrR, and SoxR, whose role as transcriptional activators in the presence of heavy metals or other stresses has been established and whose DNA-bound structures have been reported[15,38,39]. Whereas MerR proteins bind as dimers to a (pseudo-)palindromic DNA sequence and distort the DNA, which brings promoter elements into alignment for transcriptional activation (Extended Data Fig. 6i), the CarH tetramer uses its DNA-binding domains to bind to three contiguous repeat sequences in a parallel mode, which occludes the −35 element and represses transcription. This unique DNA binding mode rationalizes the tetramer architecture of two-head-to-tail dimers of CarH, which is unusual for transcription factors but here enables the DNA-binding domains to arrange in a parallel fashion and cooperatively engage the repeat sequences.

AdoCbl is a biologically expensive molecule, requiring arduous pathways for biosynthesis or specialized machinery for uptake. For CarH-using organisms, it would not be surprising if there were a recovery mechanism for Cbl following tetramer disassembly, and it is tempting to suggest that formation of bis-His ligated Cbl in some CarHs might be the first step of such a recovery pathway. In this regard, it is interesting to note that His132 is strictly conserved in thermophilic bacteria (Extended Data Fig. 3b), where perhaps the bis-His ligated Cbl is an adaption to elevated temperatures. Although use of AdoCbl as a light sensor comes at a price, it appears that the physiological benefits make this repurposing worthwhile.

Altogether, our results provide fundamental insight into a new mode of light-dependent gene regulation and reveal an exquisite example of cofactor and protein domain repurposing. The structures furthermore provide a basis for deployment of the modular CarH photoreceptors, in which the light-sensing and DNA-binding activities rest on different domains, for engineering light-modulated transcriptional control or protein–protein interactions.

Note added in proof: A paper describing a detailed photochemical mechanism for CarH based on time-resolved spectroscopic data has just been published[40].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Ziegelhoffer, E. C. & Donohue, T. J. Bacterial responses to photo-oxidative stress. *Nature Rev. Microbiol.* **7,** 856–863 (2009).
2. Eberhard, S., Finazzi, G. & Wollman, F. A. The dynamics of photosynthesis. *Annu. Rev. Genet.* **42,** 463–515 (2008).
3. Kami, C., Lorrain, S., Hornitschek, P. & Fankhauser, C. Light-regulated plant growth and development. *Plant Dev.* **91,** 29–66 (2010).
4. Palczewski, K. Chemistry and biology of vision. *J. Biol. Chem.* **287,** 1612–1619 (2012).
5. Zhang, E. E. & Kay, S. A. Clocks not winding down: unravelling circadian networks. *Nature Rev. Mol. Cell Biol.* **11,** 764–776 (2010).
6. Purcell, E. B. & Crosson, S. Photoregulation in prokaryotes. *Curr. Opin. Microbiol.* **11,** 168–178 (2008).
7. Möglich, A., Yang, X. J., Ayers, R. A. & Moffat, K. Structure and function of plant photoreceptors. *Annu. Rev. Plant Biol.* **61,** 21–47 (2010).
8. Ortiz-Guerrero, J. M., Polanco, M. C., Murillo, F. J., Padmanabhan, S. & Elias-Arnanz, M. Light-dependent gene regulation by a coenzyme B12-based photoreceptor. *Proc. Natl Acad. Sci. USA* **108,** 7565–7570 (2011).
9. Perez-Marin, M. C., Padmanabhan, S., Polanco, M. C., Murillo, F. J. & Elias-Arnanz, M. Vitamin B12 partners the CarH repressor to downregulate a photoinducible promoter in Myxococcus xanthus. *Mol. Microbiol.* **67,** 804–819 (2008).
10. Takano, H. et al. Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in Thermus thermophilus. *J. Bacteriol.* **193,** 2451–2459 (2011).

11. Elias-Arnanz, M., Padmanabhan, S. & Murillo, F. J. Light-dependent gene regulation in nonphototrophic bacteria. *Curr. Opin. Microbiol.* **14,** 128–135 (2011).

12. Takano, H. *et al.* LdrP, a cAMP receptor protein/FNR family transcriptional regulator, serves as a positive regulator for the light-inducible gene cluster in the megaplasmid of *Thermus thermophilus. Microbiology* **160,** 2650–2660 (2014).

13. Frey, P. A. in *Comprehensive Natural Products II Chemistry and Biology* Vol. 7 (eds Mander L. & Liu, H.-W.) 501–546 (Elsevier, 2010).

14. Banerjee, R. Radical carbon skeleton rearrangements: catalysis by coenzyme B_{12}-dependent mutases. *Chem. Rev.* **103,** 2083–2094 (2003).

15. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27,** 145–163 (2003).

16. Navarro-Aviles, G. *et al.* Structural basis for operator and antirepressor recognition by *Myxococcus xanthus* CarA repressor. *Mol. Microbiol.* **63,** 980–994 (2007).

17. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. How a protein binds B_{12}: a 3.0 Å X-ray structure of B_{12}-binding domains of methionine synthase. *Science* **266,** 1669–1674 (1994).

18. Diez, A. I. *et al.* Analytical ultracentrifugation studies of oligomerization and DNA-binding of TtCarH, a *Thermus thermophilus* coenzyme B_{12}-based photosensory regulator. *Eur. Biophys. J.* **42,** 463–476 (2013).

19. Cheng, Z., Li, K. R., Hammad, L. A., Karty, J. A. & Bauer, C. E. Vitamin B_{12} regulates photosystem gene expression via the CrtJ antirepressor AerR in *Rhodobacter capsulatus. Mol. Microbiol.* **91,** 649–664 (2014).

20. Marques, H. M., Marsh, J. H., Mellor, J. R. & Munro, O. Q. The coordination of imidazole and its derivatives by aquocobalamin. *Inorg. Chim. Acta* **170,** 259–269 (1990).

21. Linden, H. & Macino, G. White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa. EMBO J.* **16,** 98–109 (1997).

22. Froehlich, A. C., Liu, Y., Loros, J. J. & Dunlap, J. C. White collar-1, a circadian blue light photoreceptor, binding to the *frequency* promoter. *Science* **297,** 815–819 (2002).

23. Nash, A. I. *et al.* Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc. Natl Acad. Sci. USA* **108,** 9449–9454 (2011).

24. Burgie, E. S. & Vierstra, R. D. Phytochromes: an atomic perspective on photoactivation and signaling. *Plant Cell* **26,** 4568–4583 (2014).

25. Losi, A. & Gärtner, W. The evolution of flavin-based photoreceptors: an ancient chromophore serving trendy blue-light sensors. *Annu. Rev. Plant Biol.* **63,** 49–72 (2012).

26. Conrad, K. S., Manahan, C. C. & Crane, B. R. Photochemistry of flavoprotein light sensors. *Nature Chem. Biol.* **10,** 801–809 (2014).

27. Herrou, J. & Crosson, S. Function, structure and mechanism of bacterial photosensory LOV proteins. *Nature Rev. Microbiol.* **9,** 713–723 (2011).

28. Jenkins, G. I. The UV-B photoreceptor UVR8: from structure to physiology. *Plant Cell* **26,** 21–37 (2014).

29. Nahvi, A., Barrick, J. E. & Breaker, R. R. Coenzyme B_{12} riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32,** 143–150 (2004).

30. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9,** 1043–1049 (2002).

31. Johnson, J. E., Reyes, F. E., Polaski, J. T. & Batey, R. T. B_{12} cofactors directly stabilize an mRNA regulatory switch. *Nature* **492,** 133–137 (2012).

32. Peselis, A. & Serganov, A. Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nature Struct. Mol. Biol.* **19,** 1182–1184 (2012).

33. Jost, M., Simpson, J. H. & Drennan, C. L. The transcription factor CarH safeguards use of adenosylcobalamin as a light sensor by altering the photolysis products. *Biochemistry* **54,** 3231–3234 (2015).

34. Girvan, H. M. & Munro, A. W. Heme sensor proteins. *J. Biol. Chem.* **288,** 13194–13203 (2013).

35. Matthews, R. G. Cobalamin-dependent methyltransferases. *Acc. Chem. Res.* **34,** 681–689 (2001).

36. Jarrett, J. T. *et al.* Protein radical cage slows photolysis of methylcobalamin in methionine synthase from *Escherichia coli. Bioorg. Med. Chem.* **4,** 1237–1246 (1996).

37. Bandarian, V., Ludwig, M. L. & Matthews, R. G. Factors modulating conformational equilibria in large modular proteins: a case study with cobalamin-dependent methionine synthase. *Proc. Natl Acad. Sci. USA* **100,** 8156–8163 (2003).

38. Heldwein, E. E. Z. & Brennan, R. G. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* **409,** 378–382 (2001).

39. Watanabe, S., Kita, A., Kobayashi, K. & Miki, K. Crystal structure of the [2Fe-2S] oxidative-stress sensor SoxR bound to DNA. *Proc. Natl Acad. Sci. USA* **105,** 4121–4126 (2008).

40. Kutta, R. J. *et al.* The photochemical mechanism of a B_{12}-dependent photoreceptor protein. *Nature Commun.* **6,** 7907 (2015).

Author Contributions M.J. performed crystallographic experiments, J.F.-Z., M.C.P., and J.M.O.-G. performed *in vitro* mutant analyses, and G.K. and P.Y.-T.C. assisted with crystal structure refinement. M.J., S.P., M.E.-A., and C.L.D. designed experiments, analysed the data, and wrote the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank (PDB) under accession numbers 5C8A (AdoCbl-bound CarH crystal form 2), 5C8D (AdoCbl-bound CarH crystal form 3), 5C8E (AdoCbl- and DNA-bound CarH), and 5C8F (light-exposed CarH). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.P. (padhu@iqfr.csic.es), M.E.-A. (melias@um.es) or C.L.D. (cdrennan@mit.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein constructs. Cloning of the pET15b-CarH construct encoding for *Thermus thermophilus* CarH with an N-terminal His$_6$-tag was described previously[8]. The H132A mutation was introduced into pET15b-CarH using QuikChange PCR mutagenesis (Stratagene) with *Pfu* Turbo DNA polymerase. All other mutants were obtained by gene synthesis (Genscript) with appropriate 5′ and 3′ restriction sites for cloning into the pET15b expression vector.

Protein purification. WT CarH and mutants were purified as described previously[8]. A slightly modified protocol was used for His$_6$-tagged CarH for crystallization. After expression and affinity chromatography, performed as described previously[8], a threefold molar excess of AdoCbl (Sigma) was added and the mixture was incubated on ice for 1 h. All subsequent handling was performed in a dark room under red light. The protein solution was applied to a HiLoad 26/60 Superdex 200 size exclusion column (GE Healthcare) pre-equilibrated with CarH buffer (0.1 M NaCl, 0.05 M Tris•HCl, pH 8). Under these conditions, tetrameric AdoCbl-bound CarH eluted as a single peak, separate from residual amounts of monomeric CarH. Fractions containing AdoCbl-bound CarH were combined and concentrated to about 8 mg ml^{-1}, as judged by the absorbance at 280 nm using the combined $\varepsilon_{280\ nm}$ for AdoCbl (22.5 mM^{-1} cm^{-1}, determined spectroscopically on the basis of published extinction coefficients at 260 nm, 288 nm, and 522 nm (refs 41–43)) and for CarH (37.9 mM^{-1} cm^{-1}, calculated from the protein sequence using ProtParam at http://web.expasy.org/protparam).

Purified native and mutant protein identities were verified before use by high-performance liquid chromatography (HPLC) coupled to electrospray ionization–time-of-flight (ESI–TOF) or ion-trap mass spectrometry using an Agilent 1100 Series HPLC equipped with a µ-well plate autosampler and a capillary pump and connected to an Agilent Ion Trap XCT Plus Mass Spectrometer with an ESI interface (Agilent Technologies). Samples were injected into a Zorbax Poroshell 300 SB-C18 HPLC column (Agilent Technologies) that was coupled online to the mass spectrometer using an electrospray interface. Samples were separated at 60 °C at a flow rate of 0.2 ml min^{-1} using a linear gradient of buffer A (water/acetonitrile/formic acid, 95:4.9:0.1) to 90% buffer B (water/acetonitrile/formic acid, 10:89.9:0.1) over 30 min and protein elution was monitored at 210 nm and 280 nm. Mass spectra were acquired in the positive ion mode in an *m/z* range from 100 to 2,200.

The integrity of the AdoCbl Co–C bond was assessed by UV–vis spectroscopy (described below). Protein containing intact AdoCbl was flash-frozen in liquid nitrogen until further use. CarH containing photolysed AdoCbl was generated by exposing the protein solution to ambient light for 30 min at 4 °C. Complete photolysis was assessed by UV–vis spectroscopy (described below). Light-exposed CarH was used for crystallization experiments immediately. CarH–DNA complexes for crystallization were generated by mixing protein and DNA at the desired ratio and incubating the mixture for 1 h on ice in the dark before crystallization experiments.

Preparation of DNA segments for crystallization. HPLC-purified single-stranded DNA oligonucleotides without heavy atom labels (Integrated DNA Technologies) or containing a single 5-iodo-deoxycytidine (Jena Bioscience) were dissolved to 1 mM in CarH buffer. Equimolar amounts of complementary oligonucleotides were mixed, heated to 95 °C for 10 min, and then slowly left to cool down to 4 °C in a thermocycler over the course of 1 h for annealing. Final double-stranded DNA concentrations were assessed by the absorbance at 260 nm using the calculated sequence-specific $\varepsilon_{260\ nm}$ (http://biophysics.idtdna.com/UVSpectrum.html).

Crystallization. Purified AdoCbl-bound CarH was crystallized in three different crystal forms. All crystallization procedures for AdoCbl-bound CarH were performed in a dark room under red light. Crystals of AdoCbl-bound CarH in crystal form 1 were obtained by the hanging drop vapour diffusion technique at 25 °C. An aliquot (1 µl) of a protein solution (7 mg ml^{-1} AdoCbl-bound CarH in CarH buffer) was mixed with 1 µl of a precipitant solution (10% (w/v) PEG 8000, 10% (v/v) glycerol, 0.04 M KH$_2$PO$_4$) on a glass cover slip. The cover slip was sealed with grease over a reservoir containing 500 µl of the precipitant solution. Octahedral crystals appeared within 3 days and grew to maximum size within 7 days. Under these conditions, the protein underwent proteolysis at the linker region between the DNA-binding domain and the four-helix bundle, as judged by SDS–polyacrylamide gel electrophoresis. The crystals consisted only of the C-terminal light-sensing domains. Crystals were transferred in two steps of increasing glycerol concentration into a cryogenic solution containing 10% (w/v) PEG 8000, 20% (v/v) glycerol, 0.04 M KH$_2$PO$_4$, 0.05 M Tris•HCl pH 8, and 0.1 M NaCl, soaked in that solution for 20 s, and then flash-frozen in liquid nitrogen.

A second crystal form of AdoCbl-bound CarH was obtained by the sitting drop vapour diffusion technique at 25 °C. An aliquot (0.15 µl) of a protein solution (5.9 mg ml^{-1} AdoCbl-bound CarH in CarH buffer, supplemented with 70 µM of a 31-bp DNA oligonucleotide) was mixed with 0.15 µl of a precipitant solution (20% (w/v) PEG 3350, 0.2 M KCl) using a Phoenix liquid handling robot (Art Robbins Instruments). The drop was equilibrated against 70 µl of the precipitant solution. Rectangular crystals appeared within 6 months. Again, the protein underwent proteolysis and the crystals only consisted of the C-terminal light-sensing domains. Crystals were transferred in two steps of increasing glycerol concentration into a cryogenic solution containing the precipitant supplemented with 20% (v/v) glycerol, soaked in that solution for 5 s, and then flash-frozen in liquid nitrogen.

A third crystal form of AdoCbl-bound CarH containing full-length protein was obtained by the sitting drop vapour diffusion technique at 25 °C. An aliquot (0.15 µl) of a protein solution (6 mg ml^{-1} AdoCbl-bound CarH in CarH buffer, supplemented with 70 µM of a 28-bp DNA segment) was mixed with 0.15 µl of a precipitant solution (20% (w/v) PEG 3350, 0.1 M ammonium citrate tribasic pH 7) using a Phoenix liquid handling robot (Art Robbins Instruments). The drop was equilibrated against 70 µl of the precipitant solution. Rod crystals appeared within 1 month. These crystals contained full-length AdoCbl-bound CarH but no DNA. Crystals were transferred in three steps of increasing glycerol concentration into a cryogenic solution containing the precipitant supplemented with 20% (v/v) glycerol, soaked in that solution for 10 s, and then flash-frozen in liquid nitrogen.

Light-exposed CarH was crystallized by the hanging drop vapour diffusion technique at 25 °C. An aliquot (1 µl) of a protein solution (4.5 mg ml^{-1} light-exposed CarH in CarH buffer) was mixed with 1 µl of a precipitant solution (3.4 M NaCl, 0.1 M Bis-Tris pH 6) on a glass cover slip. The cover slip was sealed with grease over a reservoir containing 500 µl of the precipitant solution. Octahedral crystals appeared within 8 months. Crystals were transferred in three steps of increasing glycerol concentration into a cryogenic solution containing the precipitant supplemented with 18% (v/v) glycerol, incubated in that solution for 10 s, and then flash-frozen in liquid nitrogen.

CarH bound both to AdoCbl and to a 26-bp DNA segment was crystallized by the hanging drop vapour diffusion technique at 25 °C. An aliquot (1 µl) of a protein solution (6 mg ml^{-1} AdoCbl-bound CarH in CarH buffer, supplemented with 67.5 µM of a 26-bp DNA segment, 1.5-fold molar excess) was mixed with 1 µl of a precipitant solution (16% PEG 3350, 0.2 M L-proline, 0.1 M HEPES pH 7.5) on a glass cover slip. The cover slip was sealed with grease over a reservoir containing 500 µl of the precipitant solution. Tetragonal bipyramidal crystals appeared within 3 weeks. Crystals were transferred in three steps of increasing PEG 400 concentration into a cryogenic solution containing the precipitant supplemented with 15% (w/v) PEG 400, incubated in that solution for 20 s, and then flash-frozen in liquid nitrogen.

CarH bound to both AdoCbl and a 26-bp DNA segment containing 5-iodo-deoxycytidine (Extended Data Fig. 6b–d) in position −25 of the sense strand (Extended Data Fig. 5a) was crystallized by the hanging drop vapour diffusion technique at 25 °C. An aliquot (1 µl) of a protein solution (5 mg ml^{-1} AdoCbl-bound CarH in CarH buffer, supplemented with 94 µM of the iodine-labelled 26-bp DNA segment, 2.5-fold molar excess) was mixed with 1 µl of a precipitant solution (11.5% PEG 3350, 0.28 M L-proline, 0.1 M Tris pH 8.5) on a glass cover slip. The cover slip was sealed with grease over a reservoir containing 500 µl of the precipitant solution. Crystals appeared within 4 months. Crystals were transferred in five steps of increasing xylitol concentration into a cryogenic solution containing the precipitant supplemented with 25% (w/v) xylitol, incubated in that solution for 30 s, and then flash-frozen in liquid nitrogen.

Data collection and processing. All data were collected at the Advanced Photon Source (Argonne, Illinois, USA) at beamline 24ID-C using a Pilatus 6M pixel detector at a temperature of 100 K. Crystals of AdoCbl-bound CarH crystal form 1 belong to space group P4$_3$2$_1$2. An initial AdoCbl-bound CarH crystal was used for a fluorescence scan to determine the Co peak wavelength for anomalous data collection. Another crystal was then used for collection of both native and anomalous peak data. Native data were collected in a single wedge of 75° at a wavelength of 0.9792 Å (12,662 eV). The crystal was displaced continuously along its major macroscopic axis during data collection. Anomalous peak data were collected in a single wedge of 345° at a wavelength of 1.6039 Å (7,730 eV). The crystal was aligned using a mini-κ goniometer such that Bijvoet mates were recorded on the same frame.

All other data except for iodine anomalous data and native data of light-exposed CarH (see below) were collected at a wavelength of 0.9795 Å (12,658 eV). Crystals of AdoCbl-bound CarH crystal form 2 belong to space group P2$_1$2$_1$2$_1$. Data were collected in a single wedge of 100°. Crystals of AdoCbl-bound CarH crystal form 3 belong to space group P1. Data were collected in a single wedge of 270° and the crystal was displaced continuously along its major macroscopic axis during data collection. Crystals of light-exposed CarH belong to space group I4$_1$22 and data

were collected at a wavelength of 0.9791 Å (12,663 eV) in a single wedge of 150°. Crystals of DNA-bound CarH both with and without the iodine label belong to space group $P2_12_12$. Data for crystals with unlabelled DNA were collected in a single wedge of 180°. Data for crystals of CarH in complex with iodine-labelled DNA were collected at a wavelength of 1.7365 Å (7,140 eV) in a single wedge of 200° and the crystal was displaced continuously along its major macroscopic axis during data collection.

Data for the AdoCbl-bound CarH (crystal form 1) Co peak data set were integrated in HKL2000 and scaled in Scalepack[44]. Data for all other data sets were integrated in XDS and scaled in XSCALE[45]. Data collection statistics are summarized in Extended Data Table 1.

Structure building and refinement. The structure of AdoCbl-bound CarH in crystal form 1 (space group $P4_32_12$) was determined to 2.80 Å resolution using single-wavelength anomalous diffraction. Positions of two cobalt sites, corresponding to two CarH protomers in the asymmetric unit, were located using ShelxD[46] in the HKL2MAP shell[47] and refined using SHARP/autoSHARP[48]. The initial overall figure of merit (acentric) was calculated by SHARP to be 0.43 to 5.1 Å resolution. Experimental maps from the SHARP output, solvent flattened using SOLOMON[49] and extended to 3.3 Å resolution, were of sufficient quality to place two copies of the Cbl-binding domain of MetH[17] (PDB accession number 1BMT, residues 745–868), eight additional helices, and AdoCbl in the electron density. This initial model was used to better define solvent boundaries in another round of solvent flattening of SOLOMON. Using the resulting electron density, loop regions were modified and side chains with visible electron density were added. A near-complete model of AdoCbl-bound CarH (containing 374 amino-acid residues and bound AdoCbl) was then used for rigid body refinement in Phenix[50] against the native AdoCbl-bound CarH data set (crystal form 1) using data from 100 to 2.80 Å resolution. The resulting R-factors were 42.0% and 44.1% for the working and the free R-factor, respectively. The model was refined by manual adjustment in Coot[51] until rigid body refinement in Phenix yielded R-factors of 30.8% and 34.7% for the working and the free R-factor, respectively. Subsequent cycles of refinement included positional refinement with non-crystallographic symmetry restraints and individual B-factor refinement in Phenix until the R-factors were 20.9% and 24.2% for the working and the free R-factor, respectively. This model was not refined to completion. The near-complete model was used to determine the structures of AdoCbl-bound CarH in crystal form 2 (space group $P2_12_12_1$) and crystal form 3 (space group $P1$), which are of higher resolution (crystal form 2) or contain the full-length protein (crystal form 3).

The structure of AdoCbl-bound CarH in crystal form 2 was determined to 2.15 Å resolution by molecular replacement in Phaser[52]. The structure in crystal form 2 contains four CarH protomers in the asymmetric unit, corresponding to a tetramer. After molecular replacement, ten cycles of simulated annealing refinement were performed in Phenix to remove model bias. The model was then refined by iterative cycles of manual adjustment in Coot and refinement in Phenix. Initially, strict non-crystallographic symmetry restraints were applied for the two head-to-tail dimers in the asymmetric unit. Subsequently, these restraints were loosened for residues that are in unique environments either because of the asymmetric tetramer architecture or because of crystal contacts. In advanced stages of refinement, water molecules were added manually in Coot and refined in Phenix, with placement of additional water molecules until their number was stable. Final cycles of refinement included TLS parametrization[53] with one TLS group per CarH protomer.

The structure of AdoCbl-bound CarH in crystal form 3 was determined to 2.80 Å resolution using molecular replacement. First, two CarH tetramers were placed in the asymmetric unit using Phaser, accounting for all eight protomers in the asymmetric unit. Subsequently, four CarH DNA-binding domains from the structure of light-exposed CarH (see below) were placed using Phaser. After refinement in Phenix, there was clear electron density for an additional DNA-binding domain as well as fragments of the three remaining DNA-binding domains, accounting for all eight DNA-binding domains in the asymmetric unit. The model was refined by iterative cycles of manual adjustment in Coot and refinement in Phenix. Strict non-crystallographic symmetry restraints were applied for all CarH protomers in the asymmetric unit and loosened in later stages of refinement as described above. No water molecules were added to this structure. Final cycles of refinement included TLS parametrization[53]. For each CarH protomer, the light-sensing domain was defined as a single TLS group and, if fully present, the DNA-binding domain was defined as an additional TLS group.

The structure of light-exposed CarH was determined to 2.65 Å resolution by molecular replacement in Phaser using consecutive searches for the CarH Cbl-binding domain, the four-helix bundle, and the first conformation of the NMR structure of the CarA DNA-binding domain (PDB accession number 2JML[16]). The structure contains one protomer in the asymmetric unit and all three

domains could be placed unambiguously. Ten cycles of simulated annealing refinement were performed in Phenix. The model was then refined by iterative cycles of manual adjustment in Coot and refinement in Phenix. In advanced stages of refinement, water molecules were added manually in Coot and refined in Phenix, with placement of additional water molecules until their number was stable. Final cycles of refinement included TLS parametrization[53] with one TLS group.

The structure of CarH bound to AdoCbl and a 26-bp DNA segment was determined to 3.89 Å resolution by molecular replacement in Phaser using consecutive searches for two CarH tetramers without the DNA-binding domains and for two 26-bp DNA segments (models generated by the 3D-DART server[54]; http://haddock.science.uu.nl/services/3DDART/). After molecular replacement, there was clear electron density for six DNA-binding domains in the asymmetric unit, which were positioned manually in the electron density from the structure of light-exposed CarH. The model was refined by iterative cycles of manual adjustment in Coot and refinement in Phenix. B-factors were refined grouped by residue and positions of individual atoms were restrained using non-crystallographic symmetry restraints. Planarity and hydrogen bonding restraints were applied to DNA base pairs. Final cycles of refinement included TLS parametrization using one TLS group for each CarH protomer and each DNA segment. Anomalous difference maps, calculated from data collected on crystals that contained a DNA segment with an iodine label at position −25 (Extended Data Fig. 5a) were used to unambiguously determine the orientation of the DNA segment in the crystal structure and thus validate the sequence assignments. Maps, calculated using FFT[55] in the CCP4 software suite[56], revealed a strong anomalous difference density peak at one position for each of the two CarH–DNA complexes in the asymmetric unit, allowing for position −25 of the sense strand to be assigned in the structure (Extended Data Fig. 6b–d). Note that the iodine-labelled DNA segment differed slightly from the DNA segment used in the structure determination (Extended Data Fig. 5a), but both crystallize in the same space group and with the same crystal packing.

Parameter files for cobalamin were provided by O. Smart at Global Phasing. Refinement restrains for the 5′-dAdo group were generated using the Grade Web Server (Global Phasing).

Crystallographic refinement of all CarH structures yielded models possessing low free R-factors, excellent stereochemistry, and small root mean square deviations from ideal values for bond lengths and angles. In all models, side chains without visible electron density were truncated to the last atom with electron density, and amino acids without visible electron density were not included in the model. All refinement statistics are summarized in Extended Data Table 1. The models were validated using simulated annealing composite omit maps (AdoCbl-bound CarH, light-exposed CarH) or regular refinement composite omit maps (DNA-bound CarH) calculated in CNS[57] and Phenix. Model geometry was analysed using MolProbity[58] and ProCheck[59]. Analysis of the Ramachandran statistics using MolProbity indicated that for AdoCbl-bound CarH (crystal form 2), 98.1%, 1.9%, and 0.0% of residues are in the favoured, allowed, and disallowed regions, respectively; for AdoCbl-bound CarH (crystal form 3), 97.7%, 2.3%, and 0.0% of residues are in the favoured, allowed, and disallowed regions, respectively; for light-exposed CarH, 97.8%, 2.2%, and 0.0% of residues are in the favoured, allowed, and disallowed regions, respectively; and for AdoCbl- and DNA-bound CarH, 97.1%, 2.7%, and 0.2% of residues are in the favoured, allowed, and disallowed regions, respectively. The larger number of residues in the disallowed region of the Ramachandran plot of DNA-bound CarH is due to the modest resolution of the structure. Figures were generated using PyMOL[60]. Interfaces between subunits were analysed using the 'Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)[61]. Crystallography software packages were compiled by SBGrid[62].

DNA-binding assays. All DNA binding assays were repeated three to five times for each experimental condition. EMSAs were performed in the dark as described previously[8]. A 177-bp DNA probe PCR-amplified using primers with one 5′-end ^{32}P-labelled with T4 polynucleotide kinase (T4PK; Takara) before the PCR or shorter HPLC-purified synthetic probes (Biolegio) were used in the EMSAs. With the latter, one strand was ^{32}P-labelled at the 5′-end with T4PK and then mixed with a twofold excess of the unlabelled complementary strand to ensure that all of the labelled strand was present as double-stranded probe. The strand mixture was incubated at 100 °C for 2 min and then slowly left to cool down for hybridization. For EMSAs, a 20 μl reaction volume containing the DNA probe (1.2 nM, approximately 13,000 counts per minute) and protein with a fivefold excess of AdoCbl in 0.1 M KCl, 0.025 M Tris•HCl, pH 8, 1 mM DTT, 10% (v/v) glycerol, 200 ng μl⁻¹ BSA, and 1 μg of sheared salmon sperm DNA as non-specific competitor was incubated for 30 min at 65 °C (177-bp probe) or 30 °C (shorter probes). They were then loaded onto 6% native polyacrylamide gels (37.5:1 acrylamide:bisacrylamide) pre-run for

30 min in 0.5× TBE buffer (0.045 M Tris base, 0.045 M boric acid, 1 mM EDTA) and subjected to electrophoresis for 1.5 h at 200 V, 10 °C. Gels were vacuum-dried and analysed by autoradiography. Autoradiograms were scanned using an Image Scanner II imager with LabScan 5.0 software (GE Healthcare). Band intensities were quantified using ImageJ (NIH) with those of free DNA used to estimate the fraction bound, which was fitted to the three-parameter Hill equation using SigmaPlot (Systat Software) to estimate K_d, the apparent equilibrium dissociation constant equivalent to the protein concentration for half-maximal binding, and n, the Hill coefficient. The latter, for example expected to be 2 for dimer or 4 for tetramer DNA-binding models, can vary owing to cooperativity effects, contributions from monomer-tetramer equilibria, and/or deviations from true equilibrium.

DNase I and hydroxyl radical footprinting. DNase I and hydroxyl radical footprinting analyses were performed under solution conditions similar to EMSA using previously described protocols[8,63]. A 130-bp CarH operator-promoter DNA probe (Extended Data Fig. 5a) was ^{32}P-radiolabelled at the 5′ end of its sense or anti-sense strand by PCR using appropriately labelled primers, as described above. For DNase I footprinting, 20 μl of ^{32}P-radiolabelled DNA probe (~20,000 counts per minute) with 800 nM CarH and fivefold excess of AdoCbl in EMSA buffer lacking glycerol and with 0.01 M MgCl$_2$ were incubated for 30 min at 37 °C, then treated with 0.07 units of DNase I for 2 min and finally quenched with 0.025 M EDTA. For hydroxyl radical footprinting, samples (as for DNase I footprints but without MgCl$_2$) were treated with 2 μl each of freshly prepared Fe(II)-EDTA solution (1 mM ammonium iron (II) sulfate, 2 mM EDTA), 0.01 M sodium ascorbate, and 0.6% hydrogen peroxide for 4 min at 25 °C. The reaction was stopped with 2 μl each of 0.1 M thiourea and 0.5 M EDTA (pH 8). Footprinting reactions were done under dim light and, after quenching, under normal light. DNA from each sample was ethanol precipitated, washed twice with 70% ethanol, dried, and resuspended in formamide loading buffer. The 5 μl samples were heated at 95 °C for 3 min and loaded onto a 6% polyacrylamide-8 M urea sequencing gel together with G + A chemical sequencing ladders. Gels were vacuum-dried and analysed by autoradiography, and the bands quantitated using GelAnalyzer 2010a (http://www.gelanalyzer.com). Each experiment was repeated at least three times.

Analytical SEC. Analytical SEC for all CarH mutants except for H132A CarH was performed using an ÄKTAbasic unit and a Superdex 200 analytical SEC column (GE Healthcare)[8]. The calibration curve was $\log(M_r = 7.885 - 0.221 V_e)$, where M_r is the apparent molecular mass and V_e is the elution volume. Pure protein (100 μl, 50−100 μM) was incubated with a fivefold molar excess of AdoCbl for at least 15 min and analysed by SEC in the dark or after light irradiation for 5 min with white light from fluorescent lamps at 10 W m^{-2}. Elution at 0.4 ml min^{-1} flow rate was tracked by absorbance at 280 nm and 522 nm, and M_r was estimated from V_e. Each SEC experiment was performed at least three times.

Analytical SEC for H132A CarH was performed using an ÄKTA FPLC unit and a Superose 6 10/300 GL column (GE Healthcare) equilibrated with CarH buffer. The calibration curve was $\log(M_r = 9.74 - 0.30 V_e)$. WT or H132A CarH (300 μl, 20−50 μM) with stoichiometric AdoCbl with or without exposure to white light for 1 h were injected onto the column and elution was tracked by absorbance at 280 nm. For AdoCbl exchange studies, WT or H132A CarH samples were exposed to light as described and then incubated with a tenfold excess of free AdoCbl for the given periods and temperatures and analysed by SEC.

Solution UV–vis spectroscopy. Solution UV–vis spectra were recorded at 25 °C on a SpectraMax Plus 384 (Molecular Devices) using SoftMax Pro 5 software (Molecular Devices) and a 1 cm path length quartz cuvette (Starna). WT or H132A CarH in CarH buffer were transferred to the cuvette under red light or after exposure to white light for 20 min and UV–vis spectra were recorded from 250 to 800 nm. The spectrum of pure CarH buffer was used for background subtraction. No photolysis occurred on the timescale of spectrum acquisition, as repeated acquisition did not lead to spectral changes.

Spectra of Cbl with increasing imidazole concentrations, similar to spectra reported previously[20], were obtained with the same experimental parameters.

Cbl solutions contained 50 μM OHCbl•HCl (Sigma) in 50 mM Tris with 0 mM, 0.4 mM, or 400 mM imidazole, adjusted to a final pH of 8 to match the protein solutions. All solutions were incubated for 16 h at 25 °C to allow complete ligand exchange to take place.

Single crystal UV–vis spectroscopy. Single-crystal UV–vis spectra were recorded at a temperature of 100 K at Stanford Synchrotron Radiation Laboratory beamline 11-1 (Menlo Park, California, USA) using a UV–vis micro-spectrophotometer. The setup used a Hamamatsu light source (50 μm light spot) with deuterium and halogen lamps, UV solarization-resistant optical fibres, reflective Newport Schwardchild objectives, and an Ocean Optics QE65000 Spectrum Analyzer. Spectra were acquired as 50 averages with an integration time of 0.03 s and a boxcar width of 3. A crystal of AdoCbl-bound CarH was cryoprotected, transferred to a nylon fibre loop, and frozen in liquid nitrogen as described above. A background spectrum was acquired on a region of the fibre loop containing just cryoprotectant. A sample spectrum was then acquired on the crystal.

41. Barker, H. A. *et al.* Isolation and Properties of crystalline cobamide coenzymes containing benzimidazole or 5,6-dimethylbenzimidazole. *J. Biol. Chem.* **235,** 480–488 (1960).
42. Firth, R. A., Hill, H. A. O., Pratt, J. M., Williams, R. J. P. & Jackson, W. R. The circular dichroism and absorption spectra of some vitamin B$_{12}$ derivatives. *Biochemistry* **6,** 2178–2189 (1967).
43. Hill, J. A., Williams, R. J. P. & Pratt, J. M. Chemistry of vitamin B$_{12}$. Part I. The valency and spectrum of the coenzyme. *J. Chem.* Soc.5149–5153 (1964).
44. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
45. Kabsch, W. Xds. *Acta Crystallogr. D* **66,** 125–132 (2010).
46. Sheldrick, G. M. & Schneider, T. R. Substructure solution with SHELXD. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
47. Schneider, T. R. & Pape, T. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37,** 843–844 (2004).
48. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
49. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F$_1$ ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
50. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
51. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
52. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
53. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439–450 (2006).
54. van Dijk, M. & Bonvin, A. M. J. J. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* **37,** W235–W239 (2009).
55. Ten Eyck, L. F. Fast Fourier transform calculation of electron density maps. *Methods Enzymol.* **115,** 324–337 (1985).
56. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67,** 235–242 (2011).
57. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).
58. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
59. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26,** 283–291 (1993).
60. Schrodinger, L. L. C. *The PyMOL Molecular Graphics System* v.1.3r1 (Schrodinger, LLC, 2010).
61. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
62. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2,** e01456 (2013).
63. Jain, S. S. & Tullius, T. D. Footprinting protein-DNA complexes using the hydroxyl radical. *Nature Protocols* **3,** 1092–1100 (2008).
64. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).
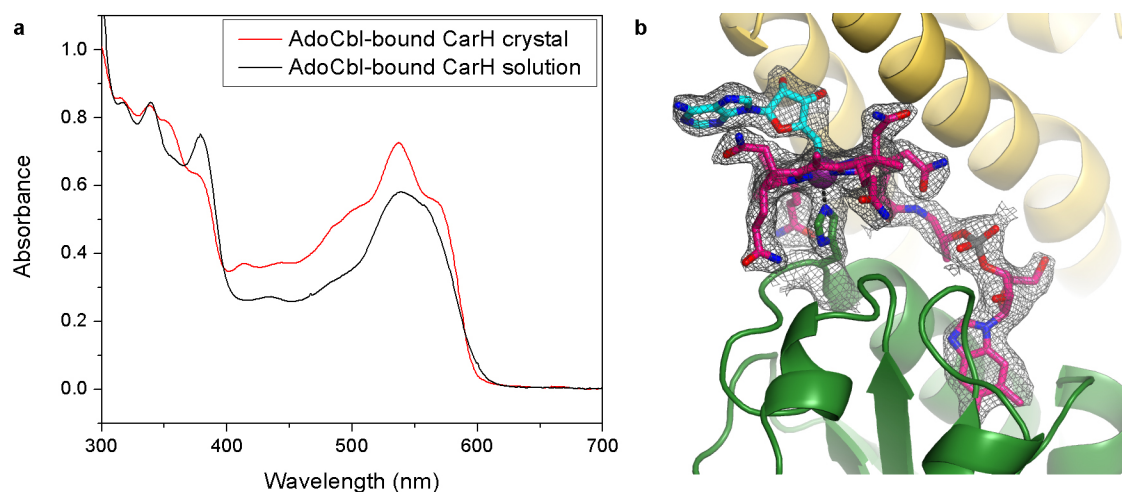
Extended Data Figure 1 | CarH crystals contain intact AdoCbl. a, UV–vis spectra obtained from AdoCbl-bound CarH crystals at $T = 100$ K (red trace) or AdoCbl-bound CarH in solution at $T = 298$ K (black trace) exhibit good qualitative agreement and similar features, including a peak centred around 540 nm with a shoulder around 560 nm. Because many band intensities are orientation-dependent and the crystal spectrum changes with orientation but molecules are rotationally averaged in solution, quantitative comparison of the spectra is difficult. Note also that individual bands appear sharper in the crystal spectrum because the molecules have fewer rotational degrees of freedom and because fewer vibrational states are populated at $T = 100$ K. **b**, Simulated annealing composite omit electron density (2.15 Å resolution) contoured around AdoCbl at 1.0σ (grey). The electron density covers the entire AdoCbl molecule including the Co–C bond, indicating that the Co–C bond remained intact during crystallization and data collection. AdoCbl is shown in stick representation with Cbl carbons in pink and 5′-dAdo group carbons in cyan. Co is shown as a purple sphere. The Co-coordinating His177 is shown in sticks with carbons in green. CarH is shown in ribbons with the helix bundle in yellow and Cbl-binding domain in green.
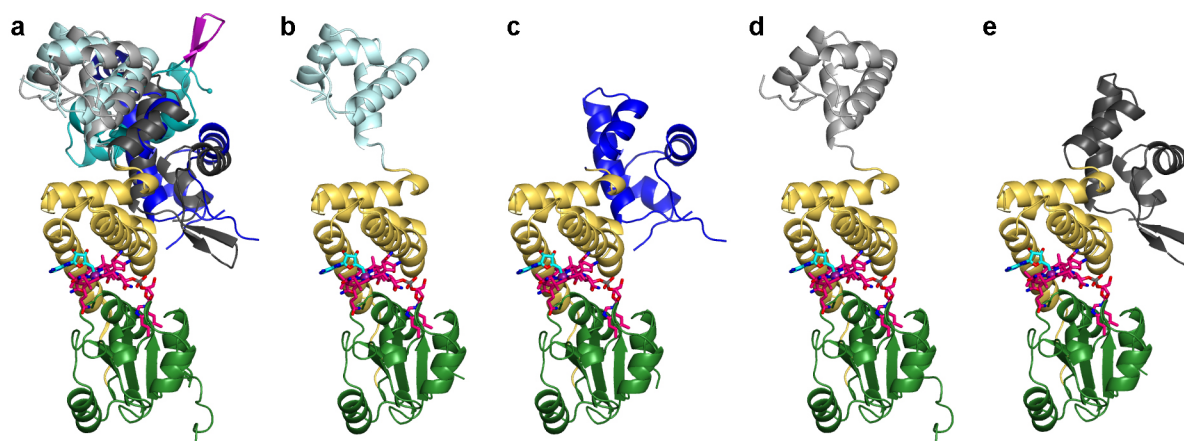
Extended Data Figure 2 | The CarH DNA-binding domain is flexible in the absence of DNA. **a**, Overlay of five CarH protomers, including the protomer shown in Fig. 2a, highlighting flexibility of DNA-binding domains. Structures are aligned by the Cbl-binding domains (green) and helix bundles (yellow) and shown in the same orientation as Fig. 2a. DNA-binding domains are coloured in dark cyan, light cyan, dark blue, black, and grey. AdoCbl is shown with Cbl carbons in pink, 5′-dAdo group carbons in cyan, and cobalt in purple. **b–e**, Individual CarH protomers shown side by side. Orientation and colouring as in **a**.

a

Mutation	Description	Oligomerization *	DNA binding *
WT		++++	++++
Q25A	DNA contact	++++	++++
R29A		++++	-
Y30A		++++	++†
H42A		++++	++†
R43A		++++	-
W131A	5'-dAdo contact	+	+
W131F		+++	+++
E141A		++	++
H142A		-	-
D201R	Dimer contact	-	-
R176D/D201R		+++	++
R176E/D201R		+++	++
G160Q	Dimer-dimer contact	++‡	++§
G192Q		++‡	++§

b

Multiple sequence alignment.

```
                              1        10        20        30        40        50        60
Thermus_thermophilus     .........MTSSGVYTIAEVEAMTGLSAEVLRQWERRYGFPKPRRTPGGHRLYSAEDVEALKTIKRWLE
Thermus_aquaticus        .........MTRPGVYTIAEVEAMTGLSAEVLRQWERRYGFPRPERTPGGHRLYRQEEVEALRTIRRWLE
Thermus_filiformis       ..........MAQDGYTIAEVEAMTGLSAEVLRQWERRYGFPRPRRTPGGHRVYSREDIEALRAIRRQLE
Meiothermus_ruber        .........MRNDLGVYTIAEVEERTGLSSALLRQWERRYGFPRPERSPGGHRLYSQTDLEALRHIKKWIA
Deinococcus_maricopensis ....MSNVSNNATGLYTTSEVEARTGVPATTLRQWERRYGLPNPTRSASGYRLYSEHDIQLIEFMQLQLA
Roseiflexus_castenholzii MAQPPSLGQLSTAPVFNTKAVARETGVPPDTFRAWERRYGIPCPQRTPGGHRLYSERDIAIIRWLRDRTA
Bacillus_megaterium      .........MAHEGKYNIKAISNMVGIQPGTLRAWERRYQILNPVRNDSGHRLYTEEDLRKLKWLTEKVS
Sorangium_cellulosum     ............MSRYRIHVVSEMCGVSSATLRAWERRYGVPSPARTASAYRLYSDDDVAVIKKMRDHVN
Myxococcus_xanthus       ..........MAERTYRINIAAELAGVRVELIRAWERRYGVLTPRRTPAGYRAYTDRDVAVLKQLKRLTD

                              70        80        90       100       110
Thermus_thermophilus     EGATPKAAIRRYLAQE..........V.....R..PEDLGTGLLEALLRGDLAGAEALFRRGLRFWGPEG
Thermus_aquaticus        EGATPKAAIRRYLAQE..........G.....S..FEGLEEGLYRALLAADLQEAEALFRRGVRLLGPEG
Thermus_filiformis       EGATPSLAIRRHLAQK..........A.....S..PEDLAPTLRQALLEGDLPQAEALFRRGVRLLGPEG
Meiothermus_ruber        EGVAPAQAVRRYLEGLT.....QEGPR.....P..PEALSMELEEALLRADTEAAERTLSEAYRLHPMET
Deinococcus_maricopensis RGITISRAAELARAR.P.....EELPA....QPATVSLAAELCAALMRADHTRAGEILSRAHAHLSVER
Roseiflexus_castenholzii EGVNISHAIMLLTSVTD.....QAPPHVEPDANRSIDRMVEDLVVALTTFDAQNAERLVSEAFAIYPFEQ
Bacillus_megaterium      GGFTISQAVSLLETESS.....TVGTFEEEGEVDSPQKIRDELLTMLLSFEEGKAQDLINHAFSLYSVEK
Sorangium_cellulosum     GGIAAAEAARMVLEAAG.PRAPVGTSGGDA.....FAGASERIVDAAVRFDPDGLEAEVNRALSLGPAVT
Myxococcus_xanthus       EGVAISEAAKLLPQLMEGLEAEVAGRGASQDARPHAETWRESMLAATQAYDQPRVSDVLDEVLAALPPLK

                             120       130       140       150       160       170       180
Thermus_thermophilus     VLEHLLLPVLREVGEAWHRGEIGVAEEHLASTFLRARLQELLDLAGF.PPGPPVLVTTPPGERHEIGAML
Thermus_aquaticus        AITRLLVPVLKRVGEAWHQGEVGVAEEHLATTFLRARLQELLDLAGL.PPGPPILVTTPPGERHEIGAML
Thermus_filiformis       AVRELLLPVLREVGEAWHQGEVSVAQEHLASTFLRARLQELLDLTGH.PYGPPILVTTPPGERHELGAMM
Meiothermus_ruber        VVMEVIAPCLRRIGDGWHLGRVSTAQEHFASTYLRGTLHGLLNLMGG.SLGPTLVVSTLPGEQHEIGSLI
Deinococcus_maricopensis VLMEVITPTLVEIGMLWERGEITVAHEHQASAYLRARLTAMLEFAGTDTWGPSVVLACAPGEQHELGLLM
Roseiflexus_castenholzii VLLEVIQPTMVEVGERWHRGEINVAAEHFATEFVRRKLASLLNVFENQPQRDTIVVGCAPQELHDLGVMF
Bacillus_megaterium      VVDIDILGSLLVTVGDMWEKGQITSAHEHYTTVQLVKTRISMIFYSLPSNGLPKAIAVCGPNETHEVGLLV
Sorangium_cellulosum     IFERALGPALERIGDLWHAGKITVAQEHLASQVLGATLLHLLRLAQPGESGRRVLLAAFADEDHVLGLFG
Myxococcus_xanthus       AFDEVLAPLLCDVGERWESGTLTVAQEHLVSQMVRARLVSLLHAAPLGR.HRHGVLACFPEEEHEMGLLG

                             190       200       210       220       230       240
Thermus_thermophilus     AAYHLRRKGVPALYLGPDTPLPDLRALARRLGAGAVVLSAVLSEPLRALP...DGALKDLAPR..VFLGG
Thermus_aquaticus        AAYALRRRGLPALYLGPDIPLPDLRSLAERLGARAVVLSVSLPDHLKALP...EGALASLGPR..VYLGG
Thermus_filiformis       AAHALRLRGLPALYLGPDTPLADLKDLAARLGARGVVLSALLPDPLRALP...DGALKGLAPR..VVLGG
Meiothermus_ruber        TALFLRRAGYTVHYLGPNTPLADLRSFAERTGAKAVVLSAVQPVSLESLP...HNALRHLAPV..VVVGG
Deinococcus_maricopensis VAVVLRRAGVRVHYMGANTPLADLAVYARLVGARAVLLSLNGTVALDETRAH.LRDLADLRMP..VFYGG
Roseiflexus_castenholzii ASLFLVRRGWHVVYLGAQVPLADLLETIRKVRPKLVCLSASTLETAMNLLEV.ARAVRQEFPHVHFGYGG
Bacillus_megaterium      FTLFLRRKGFEVIYLGSSIEDKDVELIVKEVDPTFLFMSCTMLENAEKTLNL.TNQMIKKFPHLKVGLGG
Sorangium_cellulosum     IGLRFCSWGFRSVVLGARTPPAAIARAVEALAPDLVALTVAIPPLPPSARE.LIDAYADACRGVPWIVGG
Myxococcus_xanthus       AALRLRHLGVRVTLLGQRVPAEDLGRAVLALRPDFVGLSTVASRSAEDFEDTLTRLRQALPRGLPVWVGG

                             250       260       270       280
Thermus_thermophilus     QGA..GPEEARRLG.AEYMEDLKG.L.........AEALWLPRGPEKEAI
Thermus_aquaticus        RGA..SPEEAHRLG.ARYVEGLEG.L.........AQELWQAKKEEEA..
Thermus_filiformis       AGA..DPQEARRLG.AVYLKDFSE.L.........AE.LWEREEV.....
Meiothermus_ruber        RAAASDPHLVERLG.AYYLGNDPRELAEMLSVTLKEAGLW.........
Deinococcus_maricopensis SLMNASPEVAETLG.GLFAGPDAVQAARQLLAALREAGA...........
Roseiflexus_castenholzii RIFNVNPELCAAIP.GVFLGQDARELVQTVSSIMANNGASIPPQRS....
Bacillus_megaterium      YVFDVLDSKRKGEAQPFILGNTKEE...........WNSWLTKKLAEID.
Sorangium_cellulosum     AGAGALRPFIEARGGLVAAAEPAE.IRAQVESALHA.RRRRKDGR.....
Myxococcus_xanthus       AARSHQAVCERLAVHVFQGEED...I.............WDRLAGT....
```

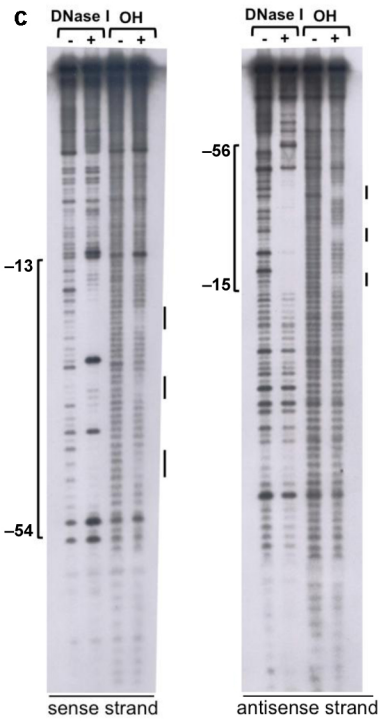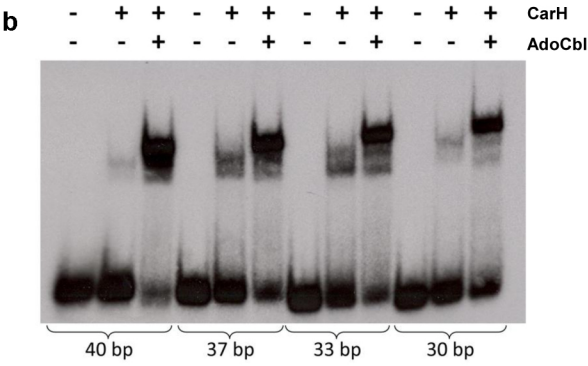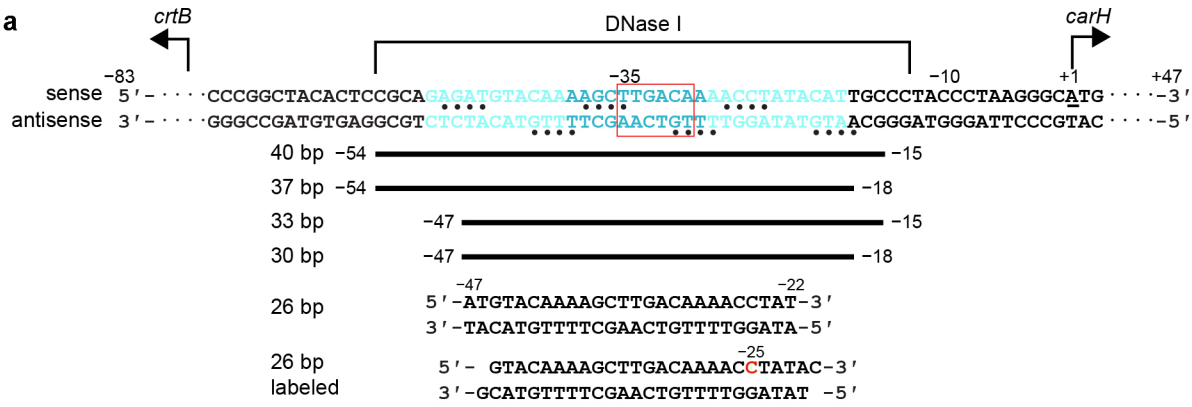**Extended Data Figure 3 | CarH mutant analysis and multiple sequence alignment. a**, Results summary for *in vitro* CarH mutant analysis. Table footnotes are as follows: *oligomerization was probed by SEC and DNA binding by gel shift analysis. †Y30A, H42A: weakened binding at 100 nM protein. ‡G160Q, G192Q: dimer, no tetramer. §G160Q, G192Q: binds with reduced affinity and cooperativity and as a higher mobility (smaller size) complex. **b**, Alignment of CarH sequences from different bacterial species. Sequence identity is shown in white font with red background, sequence similarity in red font. Coloured triangles highlight functionally important positions, with filled triangles indicating residues analysed by mutagenesis in this study and empty triangles indicating residues not analysed by mutagenesis. Mutating the highly conserved His177 of the Cbl-binding motif, the lower axial ligand of bound AdoCbl, has previously been shown to impair AdoCbl binding

and tetramerization[8]. Colouring is as follows: hydrogen bonds/ionic interactions to DNA, orange; contact to 5′-dAdo, green; histidines coordinating Cbl (His132 only coordinates after light exposure), red; hydrogen bonds/ionic interactions at dimer interface, black; hydrogen bonds/ionic interactions as well as Gly160 and Gly192 at the dimer–dimer interface, cyan. Residues involved in more than one type of interaction are coloured half/half. Residues at protein interfaces are less well conserved than other functionally important residues, probably because compensatory mutations and local structural deformations are possible. Note, however, that the *T. thermophilus* Arg176–Asp201 pair observed in our structure is inverted in *Myxococcus xanthus*, suggesting that the interaction is conserved. Alignment generated using ESPript[64].
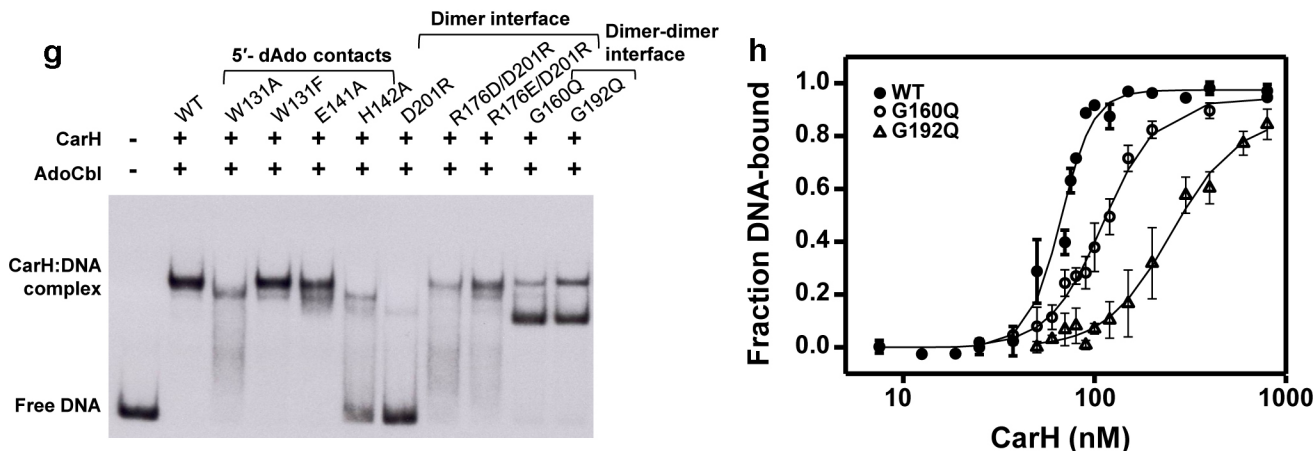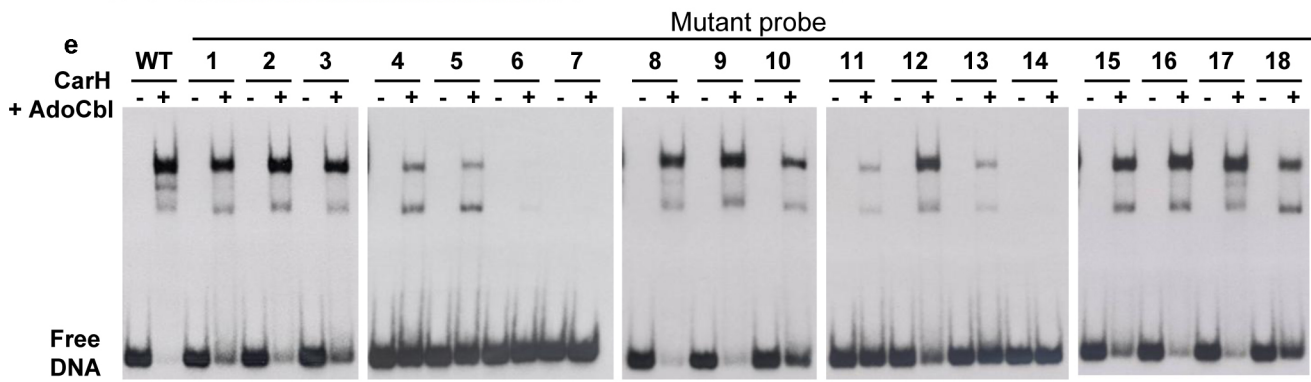
Extended Data Figure 4 | Characterization of CarH mutants affecting oligomerization state. a–f, SEC traces (Superdex 200 analytical SEC column) of CarH carrying mutations (a, b) near the 5′-dAdo group; c, d, at the head-to-tail dimer interface; and e, f, at the dimer–dimer interface. Shown are traces of mutants incubated with AdoCbl in the dark (top panels) and after light exposure (bottom panels). In all panels, both absorbance $A_{280\,nm}$ (tracking protein) and $A_{522\,nm}$ (tracking Cbl) traces are shown. Molecular masses are calculated from the observed elution volumes as described in Methods, and are consistent with a tetrameric species (137 kDa), a dimeric species (89 kDa), and a monomeric species (39 kDa). Notably, mutant CarH proteins that do not tetramerize in the presence of AdoCbl (D201R, H142A) also do not appear to bind AdoCbl (see 522 nm traces of dark samples). This finding is consistent with previous studies that show cooperativity of AdoCbl binding and tetramerization, a feature that does not hold for other forms of Cbl

(methylcobalamin, CNCbl and Cbl)[8]. Both of these mutant proteins can still bind Cbl (see 522 nm traces of light-exposed samples), which further suggests that these mutants are properly folded and that the lack of AdoCbl binding stems from inability to oligomerize. Although the degree of tetramerization of CarH mutant proteins in the dark varied, all of these mutant proteins form Cbl-bound monomers after light exposure. g, DNA-binding capacity of WT and mutant proteins (800 nM) as determined by EMSAs after incubation with AdoCbl (4 μM) in the dark. h, EMSA data for WT CarH and the G160Q and G192Q mutants fit to the Hill equation, as described in Methods. K_d (in nM) and Hill coefficients from the fits are, respectively, (67 ± 2) and (5.1 ± 0.7) for WT CarH, (111 ± 18) and (3.0 ± 0.2) for G160Q CarH, and (253 ± 17) and (2.5 ± 0.3) for G192Q CarH. The data shown are the mean values and standard errors of three to five repeat experiments.

a

b

c

d

e

Extended Data Figure 5 | Identification and validation of CarH operator sequence by EMSAs and footprinting. a, Location of CarH operator in the intergenic region between *carH* and the carotenogenic *crtB* of the *T. thermophilus* genome. Structural and biochemical data are mapped onto the sequence. Three 11-bp CarH binding sites are shown in cyan font and the promoter −35 element is highlighted with a red box. Nucleotides protected from hydroxyl radical cleavage are indicated with bullets. The ~42-nucleotide DNase I footprint on the sense strand is shown above the sequence and that of the antisense strand has been omitted for clarity. Nucleotide numbering on the sense strand is relative to the *carH* transcription start site (underlined, +1)[10]. To identify suitable DNA constructs for crystallization, operator sequences were systematically trimmed around a ~40-bp segment, as indicated by the black bars, and binding was assessed by EMSAs (shown in **b**). The sequences of two 26-bp DNA segments used for co-crystallization are also shown. The blunt-ended 26-bp segment was used for determination of the CarH–DNA structure. The second 26-bp segment contained one-nucleotide 3′-overhangs and 5-iodo-deoxycytidine in position −25 (red) and was used to validate the mode of DNA binding. **b,** Binding of CarH (800 nM) to DNA segments of different lengths after incubation with AdoCbl (4 μM) in the dark.

Substantial DNA binding was observed for a probe as small as 30-bp. **c,** DNase I and hydroxyl radical footprints of CarH on a 130-bp operator DNA segment. Disappearance of bands in the presence of CarH indicates protection from cleavage. Protected regions are marked on the side and were mapped onto the operator sequence using G + A chemical sequencing experiments performed in parallel. **d, e,** CarH binding to 40-bp operators carrying mutations. **d,** Sequences of tested operator variants. WT operator sequence shown at the top and bottom, with repeat sequences that CarH recognizes shown in cyan; 6-bp stretch contacted by CarH recognition helix is boxed. Mutations are as follows: Mut1–7: single (1–3), pairwise (4–6), and triple (7) mutations of AC to GT (positions 8/9); Mut8–14: single (8–10), double (11–13), and triple (14) mutations of (A/C)T to GC (positions 4/5); Mut15–18: pairwise (15–17) and triple (18) mutations of (A/G)A to TT (positions 1/2). **e,** EMSAs with WT CarH (800 nM) and each of the 40-bp operator variants after incubation with AdoCbl (4 μM) in the dark. Note that two additional lower mobility complexes are observed, most apparent with the WT operator and its variants with comparable binding. The origin of these complexes is unknown, but they probably arise from oligomeric equilibria and residual amounts of light-exposed protein in the sample.
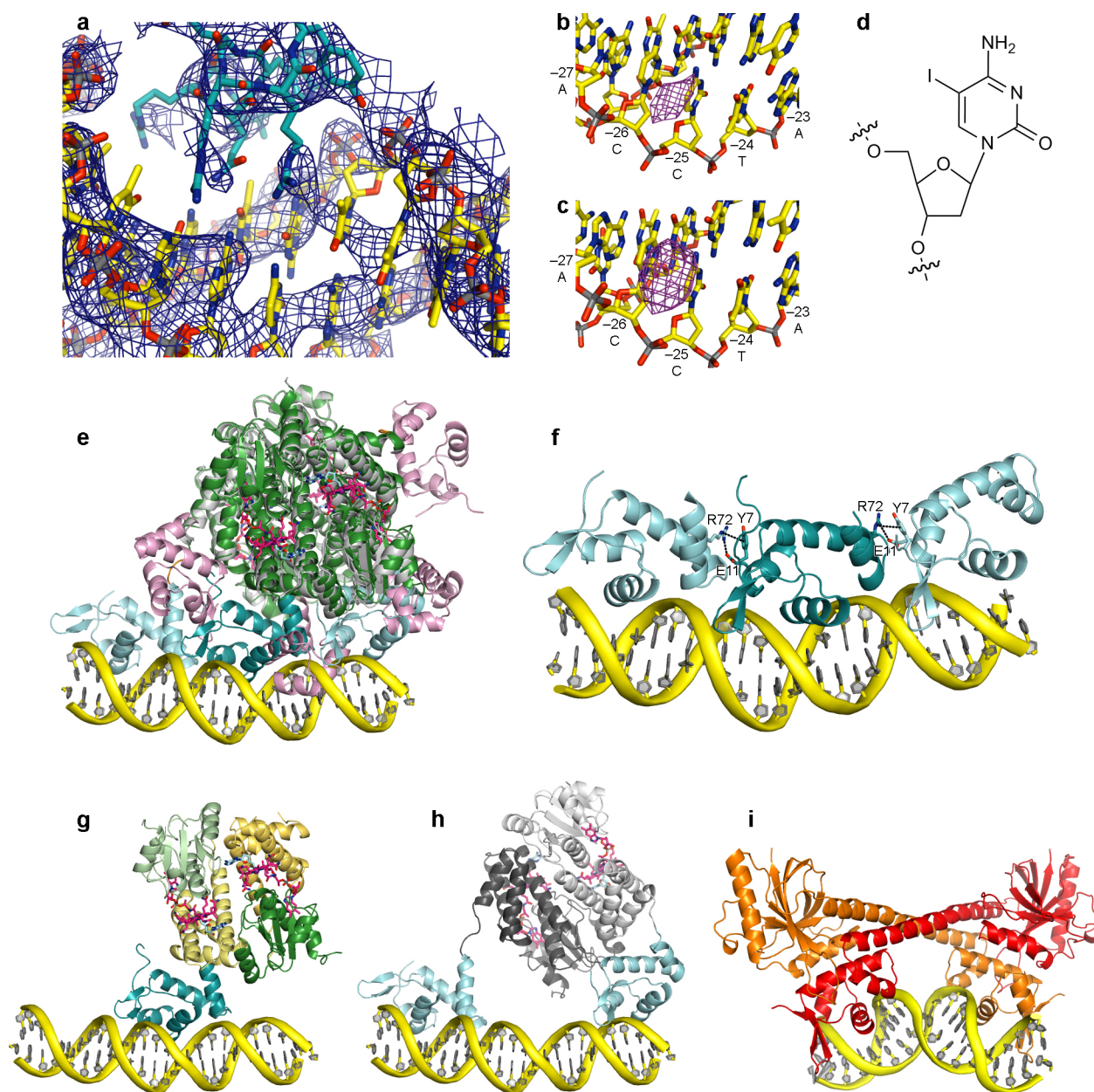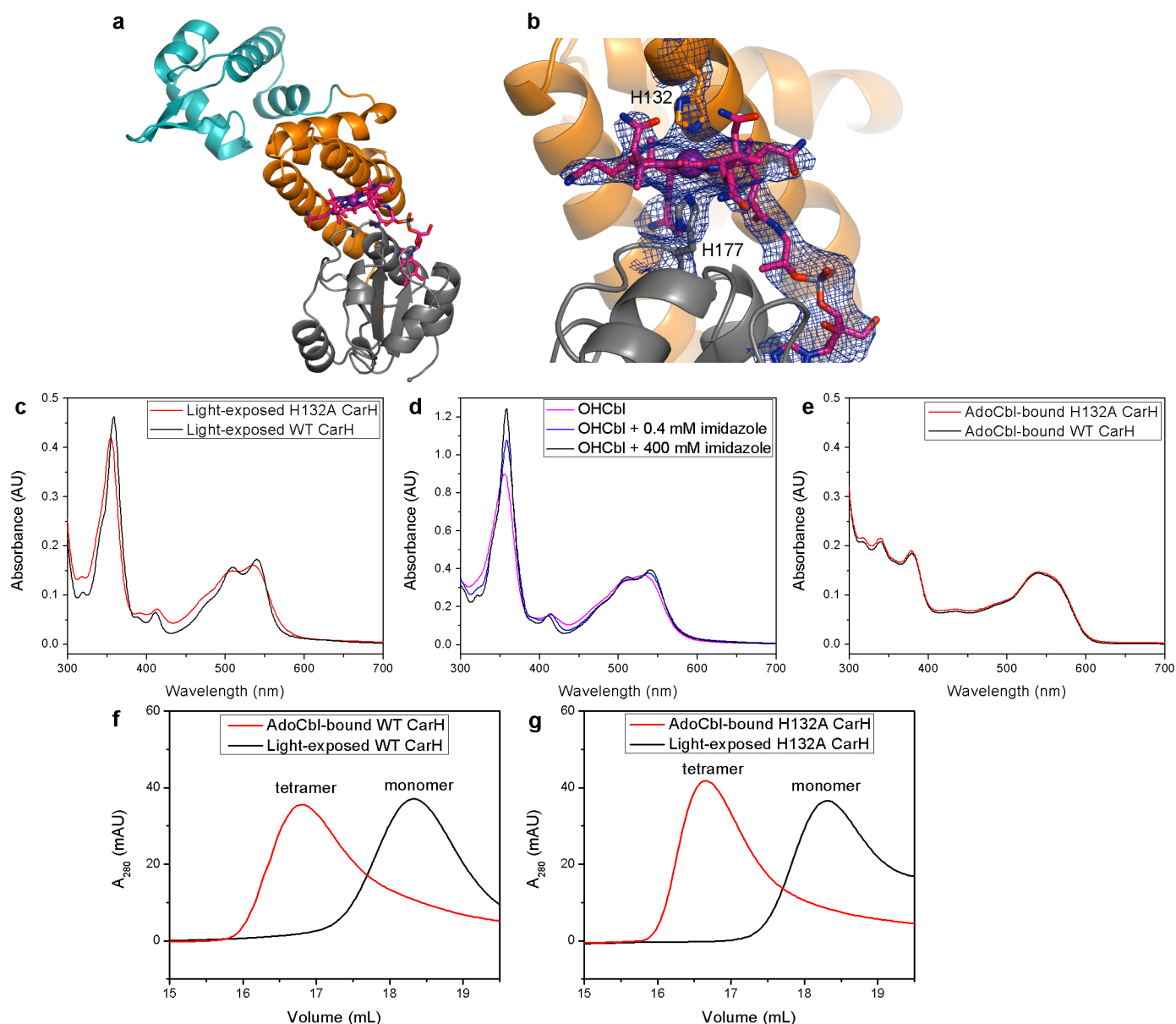
Extended Data Figure 6 | CarH DNA binding, conformational changes upon binding, and comparison with BmrR. **a**, $2F_o - F_c$ omit electron density (3.89 Å resolution) for DNA-bound CarH, calculated after performing full refinement of the model with DNA omitted and contoured at 1.0σ. DNA is shown with carbons in yellow and recognition helix of a CarH DNA-binding domain with carbons in cyan. **b, c**, Validation of DNA-binding mode using heavy-atom-derivatized DNA segments. CarH was crystallized with a DNA segment containing 5-iodo-deoxycytidine in position −25 of the sense strand. Shown is the resulting anomalous difference density (purple mesh), contoured at 6σ, for both CarH–DNA complexes in the asymmetric unit, with peaks directly adjacent to the C5 atom of deoxycytidine in position −25. **d**, Chemical structure of 5-iodo-deoxycytidine. **e**, Comparison of CarH before and after DNA binding, revealing rearrangement of DNA-binding domains. CarH before DNA binding is shown with helix bundles and Cbl-binding domains in grey and DNA-binding domains in pink. CarH bound to DNA is shown with helix bundles and Cbl-binding domains in green and DNA-binding domains in cyan. The fourth DNA-binding domain of DNA-bound CarH is disordered

and not modelled. DNA is shown in yellow. AdoCbl is shown with Cbl carbons in pink and 5′-dAdo group carbons in cyan. **f**, Contacts between residues in neighbouring DNA-binding domains, coloured by domain. Each interface between two DNA-binding domains buries 280 Å^2 of surface from solvent on each DNA-binding domain. Interactions of Arg72 to Tyr7 and Glu11 are indicated by black dashed lines. Colouring as in **e**. **g, h**, Models of individual CarH head-to-tail dimers bound to DNA. **g**, Head-to-tail dimer contributing the middle of the three DNA-binding domains, coloured by domain with DNA-binding domain in cyan, helix bundles in yellow, and Cbl-binding domains in green. The DNA-binding domain of the second protomer (right) is disordered and not modelled. DNA and AdoCbl are shown as in **e**. **h**, Head-to-tail dimer contributing the flanking DNA-binding domains. Helix bundles and Cbl-binding domains are shown in grey, remaining colouring as in **e**. **i**, BmrR bound to DNA (PDB accession number 1EXJ[38]). A BmrR dimer is shown in ribbon representation in orange and red. DNA is shown in yellow. BmrR binds as a dimer to a palindromic sequence and distorts the DNA double strand from its ideal conformation.
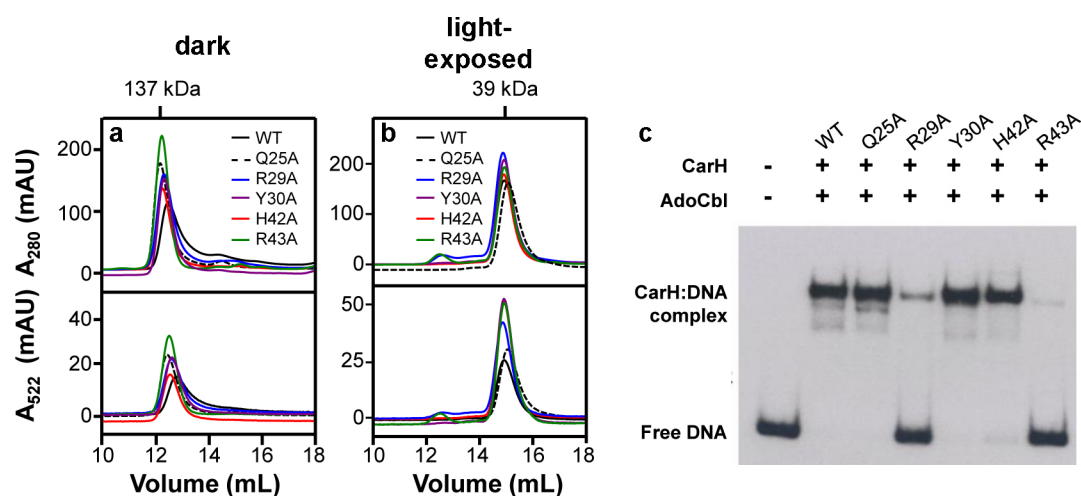
Extended Data Figure 7 | *In vitro* characterization of CarH DNA binding mutants. **a**, **b**, SEC traces (Superdex 200 analytical SEC column) of CarH carrying mutations in the DNA-binding domain. Shown are traces of mutants (**a**) incubated with AdoCbl in the dark and (**b**) after light exposure. In all panels, both $A_{280\ nm}$ (tracking protein) and $A_{522\ nm}$ (tracking Cbl) traces are shown. Molecular masses are calculated from the observed elution volumes as described in Methods, and are consistent with a tetrameric species (137 kDa) and a monomeric species (39 kDa). **c**, DNA-binding capacity of mutants (800 nM) as determined by EMSAs after incubation with AdoCbl (4 μM) in the dark.

Extended Data Figure 8 | Light-exposed CarH has bis-His ligated Cbl.
a, Structure of light-exposed CarH including the DNA-binding domain (cyan) and other domains coloured as in Fig. 5a. **b**, Close-up view of the Cbl in light-exposed CarH, with both coordinating His side chains shown in sticks. Simulated annealing composite omit electron density (2.65 Å resolution) is shown in blue, contoured at 1.0σ. (**c**) UV–vis spectra of light-exposed WT CarH (black) and H132A CarH (red) exhibit pronounced differences, indicating that the bis-His ligation is also formed in solution. **d**, UV–vis spectra of free OHCbl (50 μM) with increasing imidazole concentration. The spectrum of bis-imidazole ligated Cbl (black, Cbl with 400 mM imidazole contains 60% bis-imidazole ligated Cbl and 40% Cbl with dimethylbenzimidazole and imidazole as ligands[20]) resembles that of light-exposed WT CarH, whereas the spectrum of free OHCbl (pink) resembles that of light-exposed H132A CarH. Note that the latter two are expected to be slightly different because free OHCbl contains a dimethylbenzimidazole group as the lower axial ligand, whereas light-exposed H132A CarH contains a histidine imidazole as the lower axial ligand. Experimental conditions chosen were similar to those reported elsewhere[20]. **e**, UV–vis spectra of AdoCbl-bound WT CarH (black) and H132A CarH (red) are virtually identical, suggesting that the mode of AdoCbl binding is unchanged, as is expected from the structure. **f**, **g**, Size-exclusion chromatograms (Superose 6 10/300 GL column) of AdoCbl-bound and light-exposed (**f**) WT CarH and (**g**) H132A CarH, demonstrating that H132A CarH, like WT CarH, forms a tetramer in the dark and undergoes light-dependent tetramer disassembly.
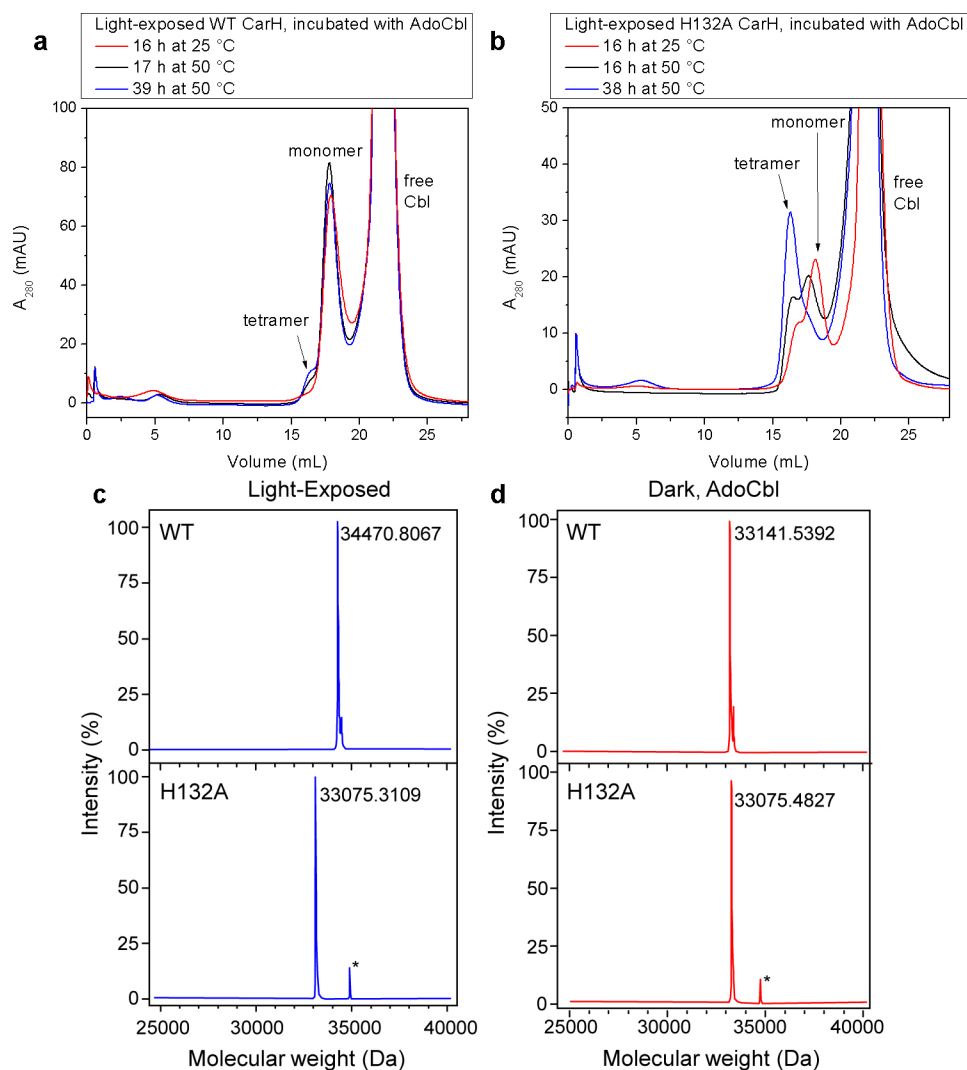
Extended Data Figure 9 | Disruption of bis-His ligation by H132A mutation facilitates Cbl dissociation after photolysis. a, b, WT and H132A CarH were exposed to light, rendering them monomeric, and then incubated with free AdoCbl at the indicated temperatures and periods. For AdoCbl to bind to CarH and induce tetramerization, the photolysed Cbl has to dissociate from the protein first. Thus, the extent of tetramer formation, as assessed by SEC (Superose 6 10/300 GL column), is indicative of the affinity of the protein for photolysed Cbl. That is, lack of tetramer formation in the presence of fresh AdoCbl indicates that the photolysed Cbl is still bound to the protein. The observed differential in tetramer formation between H132A and WT CarH is substantial: WT CarH retains its photolysed Cbl, showing only a small amount of tetramer formation, whereas H132A CarH loses its photolysed Cbl, reforming tetramers upon AdoCbl addition. **c,** ESI–TOF mass spectra of WT and H132A CarH after light exposure also reveal differential affinity for photolysed Cbl. Light-exposed WT CarH is 1,329 Da larger in molecular mass than light-exposed H132A CarH, a number corresponding to the molecular mass of Cbl. This difference in molecular mass suggests that, even under the harsh conditions of this experiment, photolysed Cbl remains bound to the WT CarH monomer but not to H132A CarH, indicating that Cbl dissociates more readily without the bis-His ligation. The minor peak next to WT CarH arises from protein bound to a potassium ion (mass shift 39 Da). Species marked with an asterisk correspond to an unidentified impurity in the H132A CarH sample. **d,** Control experiment showing ESI–TOF mass spectra of WT and H132A CarH in the AdoCbl-bound dark state. The very similar molecular masses obtained (differing only because of the H132A mutation) indicate that the mutation has no effect on AdoCbl binding, consistent with the fact that His132 is not coordinated to Cbl when the upper 5′-dAdo ligand is present. Both WT and H132A CarH lose their AdoCbl cofactor as the tetramer disassembles into monomeric units.

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	AdoCbl-bound CarH form 1 Native *	AdoCbl-bound CarH form 1 Co Peak †	AdoCbl-bound CarH form 2	AdoCbl-bound CarH form 3	AdoCbl- and DNA-bound CarH	AdoCbl- and DNA-bound CarH (iodine-labeled) *†	light-exposed CarH
PDB code			5C8A	5C8D	5C8E		5C8F
Data collection							
Space group	$P4_32_12$	$P4_32_12$	$P2_12_12_1$	$P1$	$P2_12_12$	$P2_12_12$	$I4_122$
Cell dimensions							
a, b, c (Å)	94.5, 94.5, 180.5	94.5, 94.5, 180.6	51.4, 99.7, 144.0	78.7, 79.7, 118.4	177.9, 141.8, 162.7	176.7, 141.7, 162.6	126.9, 126.9, 149.5
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.7, 96.6, 117.3	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.9792	1.6039	0.9795	0.9795	0.9795	1.7365	0.9791
Resolution (Å)	200 – 2.80 (2.87 – 2.80)	200 – 3.30 (3.36 – 3.30)	100 – 2.15 (2.21 – 2.15)	100 – 2.80 (2.87 – 2.80)	100 – 3.89 (3.99 – 3.89)	100 – 5.00 (5.13 – 5.00)	100 – 2.65 (2.72 – 2.65)
R_{sym} (%) ‡	6.5 (69.0)	12.4 (38.6)	5.3 (60.8)	10.5 (83.2)	9.8 (126.4)	7.2 (107.1)	6.9 (167.8)
R_{meas} (%) ‡	7.2 (76.5)	– §	6.2 (71.5)	11.7 (91.9)	10.2 (131.9)	8.6 (127.2)	7.2 (174.4)
$CC_{1/2}$ ‡	99.9 (77.1)	– §	99.9 (87.7)	99.7 (84.3)	99.9 (87.0)	99.9 (54.3)	100.0 (68.5)
$< I / \sigma(I) >$ ‡	16.4 (2.3)	15.7 (7.0)	15.2 (2.1)	12.0 (2.0)	15.7 (2.0)	7.4 (1.1)	26.8 (1.8)
Completeness (%)‡	99.5 (99.9)	99.5 (100.0)	98.5 (99.4)	94.9 (95.5)	99.9 (100.0)	97.7 (98.2)	100.0 (99.9)
Redundancy‡	5.3 (5.4)	11.9 (11.5)	3.6 (3.6)	5.6 (5.6)	12.2 (12.4)	3.5 (3.3)	12.8 (13.4)
Refinement							
Resolution (Å) ‡			100 – 2.15 (2.21 – 2.15)	100 – 2.80 (2.87 – 2.80)	100 – 3.89 (3.99 – 3.89)		100 – 2.65 (2.72 – 2.65)
No. reflections‡			40511 (2970)	59284 (4419)	38480 (2819)		18067 (1319)
R_{work}/ R_{free}			0.183/0.227	0.183/0.230	0.250/0.257		0.172/0.203
No. atoms							
protein			5766	14668	14500		2090
Cbl			364	728	728		91
5′-deoxyadenosine			72	144	144		–
water			259	–	–		19
DNA			–	–	2120		–
glycerol			6	–	–		6
chloride			–	–	–		2
B-factors							
protein			48.5	76.4	164.4		85.9
Cbl			45.1	65.2	153.9		100.8
5′-deoxyadenosine			46.1	75.8	158.5		–
water			47.3	–	–		71.2
DNA			–	–	231.3		–
glycerol			54.3	–	–		95.3
chloride			–	–	–		94.7
R.m.s deviations							
Bond lengths (Å)			0.004	0.005	0.005		0.004
Bond angles (°)			0.82	0.92	0.79		0.74
Rotamer outliers (%)			5 (0.9%)	8 (0.6%)	10 (0.8%)		0 (0.0%)

* Structure was not refined to completion.
† Bijvoet pairs were not merged during data processing.
‡ Values in parentheses indicate highest-resolution bin.
§ Values were not reported in the version of Scalepack used for scaling.

LETTER

Flows of X–ray gas reveal the disruption of a star by a massive black hole

Jon M. Miller[1], Jelle S. Kaastra[2,3,4], M. Coleman Miller[5], Mark T. Reynolds[1], Gregory Brown[6], S. Bradley Cenko[7,8], Jeremy J. Drake[9], Suvi Gezari[5], James Guillochon[10], Kayhan Gultekin[1], Jimmy Irwin[11], Andrew Levan[6], Dipankar Maitra[12], W. Peter Maksym[11], Richard Mushotzky[5], Paul O'Brien[13], Frits Paerels[14], Jelle de Plaa[2], Enrico Ramirez-Ruiz[15], Tod Strohmayer[7] & Nial Tanvir[13]

Tidal forces close to massive black holes can violently disrupt stars that make a close approach. These extreme events are discovered via bright X-ray[1–4] and optical/ultraviolet[5,6] flares in galactic centres. Prior studies based on modelling decaying flux trends have been able to estimate broad properties, such as the mass accretion rate[6,7]. Here we report the detection of flows of hot, ionized gas in high-resolution X-ray spectra of a nearby tidal disruption event, ASASSN-14li in the galaxy PGC 043234. Variability within the absorption-dominated spectra indicates that the gas is relatively close to the black hole. Narrow linewidths indicate that the gas does not stretch over a large range of radii, giving a low volume filling factor. Modest outflow speeds of a few hundred kilometres per second are observed; these are below the escape speed from the radius set by variability. The gas flow is consistent with a rotating wind from the inner, super-Eddington region of a nascent accretion disk, or with a filament of disrupted stellar gas near to the apocentre of an elliptical orbit. Flows of this sort are predicted by fundamental analytical theory[8] and more recent numerical simulations[7,9–14].

ASASSN-14li was discovered in images obtained on 22 November 2014 (modified Julian day MJD 56,983), at a visual magnitude of $V = 16.5$ (ref. 15) by the All-Sky Automated Survey for Supernovae (ASASSN). Follow-up observations found this transient source to coincide with the centre of the galaxy PGC 043234 (originally Zwicky VIII 211), to within 0.04 arcseconds (ref. 15). This galaxy lies at a redshift of $z = 0.0206$, or a luminosity distance of 90.3 Mpc (for $H_0 = 73 \text{ km s}^{-1}$, $\Omega_{matter} = 0.27$, $\Omega_\Lambda = 0.73$), making ASASSN-14li the closest disruption event discovered in over ten years. The discovery magnitudes indicated a substantial flux increase over prior, archival optical images of this galaxy. Follow-up observations with the Swift space observatory's X-ray Telescope[16,17] (XRT) established a new X-ray source at this location[15].

Archival X-ray studies rule out the possibility that PGC 043234 harbours a standard active galactic nucleus that could produce bright flaring. PGC 043234 is not detected in the ROSAT All-Sky Survey[18]. Using the online interface to the data, the background count rate for sources detected in the vicinity is 0.002 counts s^{-1} arcmin^{-2}. With standard assumptions (see Methods), this rate corresponds to a luminosity of $L \approx 4.8 \times 10^{40} \text{ erg s}^{-1}$, which is orders of magnitude below a standard active nucleus.

Theory predicts that early tidal disruption event (TDE) evolution should be dominated by a bright, super-Eddington accretion phase, and be followed by a characteristic $t^{-5/3}$ decline as disrupted material

interacts and accretes[8,19]. Detections of winds integral to super-Eddington accretion have not been reported previously, but $t^{-5/3}$ flux decay trends in the ultraviolet part of the spectrum (where disk emission from active nuclei typically peaks) are now a standard signature of TDEs in the literature[5,6]. Figure 1 shows the flux decay of ASASSN-14li, as observed by Swift. A fit to the UVM2-filter data assuming an index of $\alpha = -5/3$ gives a disruption date of $t_0 \approx 56{,}948 \pm 3$ (MJD). The V-band light is consistent with a shallower $t^{-5/12}$ decay; this can indicate direct thermal emission from the disk, or reprocessed emission[7,11] (see Methods).

We triggered approved XMM-Newton programs to study ASASSN-14li soon after its discovery. Although the space observatory XMM-Newton carries several instruments, the spectra from the two Reflection Grating Spectrometer (RGS) units are the focus of this analysis. We were also granted a Director's Discretionary Time
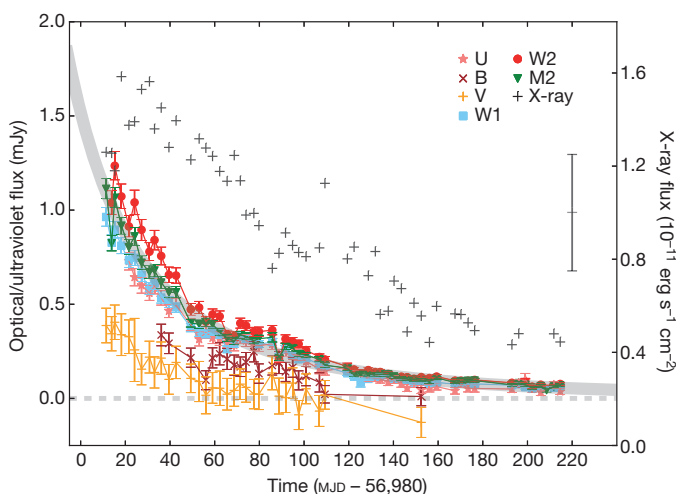

Figure 1 | The multi-wavelength light curves of ASASSN-14li clearly signal a tidal disruption event. The light curves are based on monitoring observations with the Swift satellite. The errors shown on plotting symbols are the 1σ confidence limits on the flux in each band (V, B, U, UVW1 (here W1), UVM2 (M2), UVW2 (W2)). Contributions from the host galaxy have been subtracted (see Methods). The UVM2 filter samples the ultraviolet light especially well. The grey shading depicts the $t^{-5/3}$ flux decay predicted by fundamental theory[8,19]. The X-ray flux points carry relatively large errors; a representative error bar is shown at right. Fits to the decay curve are described in the main text and in the Methods.

[1]Department of Astronomy, The University of Michigan, 1085 South University Avenue, Ann Arbor, Michigan 48103, USA. [2]SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands. [3]Department of Physics and Astronomy, Universiteit Utrecht, PO Box 80000, 3508 TA Utrecht, The Netherlands. [4]Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands. [5]Department of Astronomy, The University of Maryland, College Park, Maryland 20742, USA. [6]Department of Physics, University of Warwick, Coventry CV4 7AL, UK. [7]Joint Space-Science Institute, University of Maryland, College Park, Maryland 02742, USA. [8]Astrophysics Science Division, NASA Goddard Space Flight Center, MC 661, Greenbelt, Maryland 20771, USA. [9]Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [10]The Institute for Theory and Computation, Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [11]Department of Physics and Astronomy, University of Alabama, PO Box 870324, Tuscaloosa, Alabama 35487, USA. [12]Department of Physics and Astronomy, Wheaton College, Norton, Massachusetts 02766, USA. [13]Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK. [14]Columbia Astrophysics Laboratory and Department of Astronomy, Columbia University, 550 West 120th Street, New York, New York 10027, USA. [15]Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA.

Table 1 | Modelling of the high-resolution X-ray spectra reveals ionized flows of gas

Mission	XMM-Newton	XMM-Newton	XMM-Newton	XMM-Newton	Chandra	XMM-Newton
Observation ID	0694651201	0722480201	0722480201	0722480201	17566, 17567	0694651401
Comment	Monitoring	Long stare	Stare (low)	Stare (high)	None	Monitoring
Start (MJD)	56,997.98	56,999.54	56,999.94	57,000.0	56,999.97, 57,002.98	57,023.52
Duration (ks)	22	94	36	58	35, 45	23.6
$F_{X,b}$ (10^{-11} erg cm^{-2} s^{-1})	2.7 ± 0.7	3.2 ± 0.4	3.4 ± 0.3	3.4 ± 0.2	$2.5^{+0.2}_{-0.3}$	2.68 ± 0.08
$L_{X,b}$ (10^{44} erg s^{-1})	2.9 ± 0.7	2.2 ± 0.3	2.2 ± 0.2	2.0 ± 0.1	$1.7^{+0.1}_{-0.2}$	3.2 ± 0.1
$F_{X,f}$ (10^{-11} erg cm^{-2} s^{-1})	1.2 ± 0.3	1.2 ± 0.2	1.07 ± 0.08	1.24 ± 0.08	$1.0^{+0.1}_{-0.2}$	1.19 ± 0.04
$L_{X,f}$ (10^{44} erg s^{-1})	0.25 ± 0.06	0.21 ± 0.03	0.19 ± 0.01	0.21 ± 0.01	$0.17^{+0.01}_{-0.02}$	0.27 ± 0.01
$N_{H,MW}$ (10^{20} cm^{-2})	2.6*	2.6 ± 0.6	2.6*	2.6*	2.6*	2.6*
$N_{H,HG}$ (10^{20} cm^{-2})	1.4*	1.4 ± 0.5	1.4*	1.4*	1.4*	1.4*
$N_{H,TDE}$ (10^{22} cm^{-2})	0.7 ± 0.2	$1.3^{+0.9}_{-0.4}$	$0.1^{+0.3}_{-0.2}$	$0.9^{+0.2}_{-0.3}$	$0.5^{+0.4}_{-0.1}$	0.5 ± 0.1
$\log \xi$ (erg cm s^{-1})	3.6 ± 0.1	4.1 ± 0.2	4.1 ± 0.1	$3.9^{+0.3}_{-0.1}$	$3.9^{+0.1}_{-0.2}$	3.7 ± 0.1
v_{rms} (km s^{-1})	130 ± 30	110^{+30}_{-20}	60^{+60}_{-50}	120 ± 20	120^{+40}_{-30}	230^{+60}_{-50}
v_{shift} (km s^{-1})	-180 ± 60	-210 ± 40	-360 ± 50	-130^{-50}_{+70}	-500^{+60}_{-70}	-490 ± 70
kT (eV)	50.0 ± 0.09	51.4 ± 0.1	50.0 ± 0.4	52.6 ± 0.4	52.6 ± 0.3	49.7 ± 0.9
Emitting area (10^{25} cm^2)	5.7 ± 1.4	3.7 ± 0.5	4.0 ± 0.3	3.0 ± 0.2	$2.5^{+0.1}_{-0.2}$	6.1 ± 0.2
χ^2/ν	704.8/567	870.5/563	687.8/564	726.8/565	266.5/178	626.5/566

Each spectrum was fitted with a simple blackbody continuum, modified by photoionized absorption via the *pion* model, and interstellar absorption in the host galaxy PGC 043234 and the Milky Way. The fits were made using SPEX[28], minimizing a χ^2 statistic. In all cases, 1σ errors are quoted. Where a parameter is quoted with an asterisk, the listed parameter was not varied. X-ray fluxes F_X and luminosities L_X listed with the subscript 'b' for 'broad' were extrapolated from the fitting band to the 1.24–124 Å band; those with the subscript 'f' represent values for the 18–35 Å fitting band. Interstellar column densities N_H are separately measured for the Milky Way ($N_{H,MW}$) at zero redshift and the host galaxy PGC 043234 ($N_{H,HG}$) at a redshift of $z = 0.0206$. These parameters were measured in the XMM-Newton 'long stare' and then fixed in fits to other spectra. Variable parameters in the photoionization model are listed together; the negative v_{shift} values indicate a blueshift relative to the host galaxy. Here ξ is the ionization parameter of the gas, v_{rms} is the root-mean-square velocity width of the spectral lines, k is Boltzmann's constant, T is temperature and ν is the number of degrees of freedom.

observation with the Chandra X-ray Observatory, using its Low Energy Transmission Grating spectrometer (LETG), paired with its High Resolution Camera for spectroscopy (HRC-S).

The 18–35 Å X-ray spectra of ASASSN-14li are clearly thermal in origin, so we modelled the continuum with a single blackbody, modified by interstellar absorption in PGC 043234 and the Milky Way, and absorption from blueshifted, ionized gas local to the TDE. The self-consistent photoionization code *pion*[20] was used to model the complex absorption spectra (see Table 1 and Methods).

Assuming that the highest bolometric luminosity derived from fits to the high-resolution spectra ($L = 3.2 \pm 0.1 \times 10^{44}$ erg s^{-1}) corresponds to the Eddington limit, a black-hole mass of $2.5 \times 10^6 M_\odot$,

where $M_\odot$ is the mass of the Sun, is inferred. The blackbody emission measured from fits to the time-averaged XMM-Newton spectrum gives an emitting area of 3.7×10^{25} cm^2, implying $r = 1.7 \times 10^{12}$ cm for a spherical geometry.

This is consistent with the innermost stable circular orbit around a black hole of mass $M \approx 1.9 \times 10^6 M_\odot$. Modelling of the Swift light curves (see Fig. 1), using a self-consistent treatment of direct and reprocessed light from an elliptical accretion disk[7] gives a mass in the range of $M \approx (0.4$–$1.2) \times 10^6 M_\odot$ (see Methods). Together, the thermal spectrum, implied radii and the run of emission from X-rays to optical bands unambiguously signal the presence of an accretion disk in ASASSN-14li.
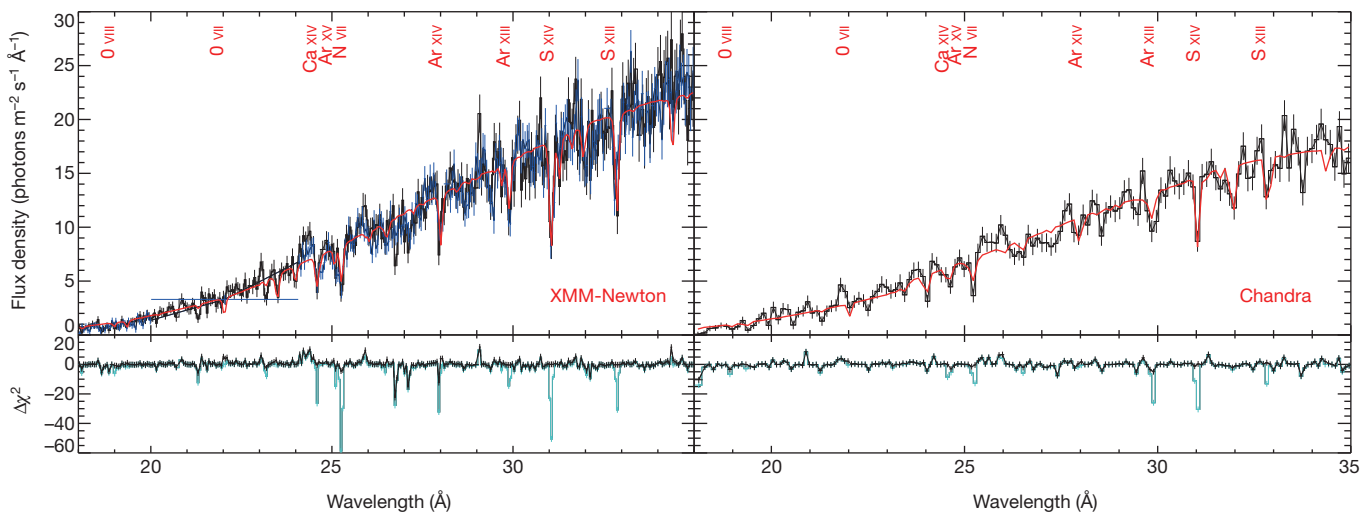

Figure 2 | The high-resolution X-ray spectra of ASASSN-14li reveal blueshifted absorption lines. Spectra from the 'long stare' with XMM-Newton and the combined Chandra spectrum are shown. XMM-Newton spectra from the RGS1 and RGS2 units are shown in black and blue, respectively; the RGS2 unit is missing a detector in the 20–24 Å band. The best-fit photoionized absorption model for the outflowing gas detected in each spectrum is shown in red (see Methods), and selected strong lines are indicated. Below each spectrum, the goodness-of-fit statistic ($\Delta \chi^2$) is shown before (cyan) and after (black) modelling the absorbing gas. The errors on the spectra are 1σ confidence limits on the flux in each bin.
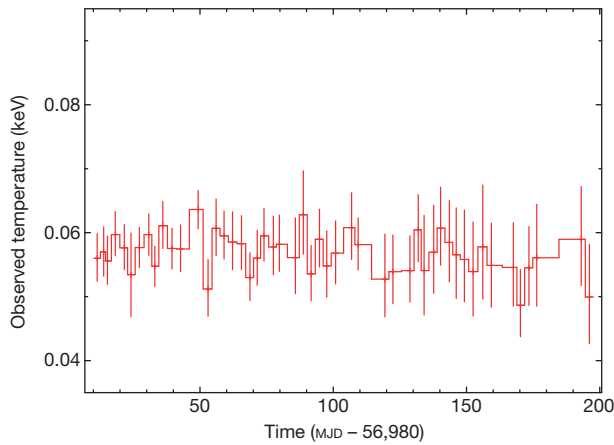
Figure 3 | The temperature of the blackbody continuum emission from ASASSN-14li is steady over time. The temperature measured in simple blackbody fits to Swift/XRT monitoring observations is plotted versus time. Errors are 1σ confidence intervals. The temperature is remarkably steady, contrasting strongly with the declining fluxes shown in Fig. 1. Recent theory suggests that winds may serve to maintain steady temperatures in some TDEs[14].

Figure 2 shows the best-fit model for the spectra obtained in the 'long stare' with the XMM-Newton/RGS (see Table 1 and Methods). An F-test finds that photoionized X-ray absorption is required in fits to these spectra at more than the 27σ level of confidence, relative to a spectral model with no such absorption. The model captures the majority of the strong absorption lines, giving $\chi^2 = 870.5$ for 563 degrees of freedom (see Table 1). The strongest lines in the spectrum coincide with ionized charge states of N, O, S, Ar and Ca. Only solar abundances are required to describe the spectra. The Chandra spectrum independently confirms these results in broad terms, and requires absorption at more than the 6σ level of confidence.

A hard lower limit on the radius of the absorbing gas is set by the blackbody continuum. The best radius estimate probably comes from variability timescales within the XMM-Newton 'long stare'. Analysis of specific time segments within the 'long stare', as well as flux-selected segments, reveals that the absorption varies (see Table 1 and Methods). This sets a relevant limit of $r \lesssim c\delta t$, or $r \lesssim 3 \times 10^{15}$ cm, where r is the radius of the absorbing gas relative to the central engine, c is the speed of light, and δt is the time interval of the variability. Although the column density and ionization do not vary significantly, the blueshift of the gas does. During the initial third of the observation, the blueshift is larger, $v_{shift} = -360 \pm 50$ km s^{-1}, but falls to $v_{shift} = -130^{-50}_{+70}$ km s^{-1} in the final two-thirds. Shorter monitoring observations with XMM-Newton reveal evolution of the absorbing gas, including changes in ionization and column density, before and after the 'long stare' (see Table 1 and Methods).

Fundamental theoretical treatments of TDEs predict an initial near-Eddington or super-Eddington phase[8]; this is confirmed in more recent theoretical studies[9,12,21]. The high-resolution X-ray spectra were obtained within the predicted time frame for super-Eddington accretion, for our estimates of the black-hole mass[22]. Although the ionization parameter of the observed gas is high, the ionizing photon distribution peaks at a low energy, and the wind could be driven by radiation force. Such flows are naturally clumpy, and may be similar to the photospheres of novae[23]. Given the strong evidence of an accretion disk in our observations of ASASSN-14li, the X-ray outflow is best associated with a wind from the inner regions of a nascent, super-Eddington accretion disk. The local escape speed at an absorption radius of $r \approx 10^4 \, GM/c^2$ (appropriate for $M \approx 10^6 M_\odot$; G is Newton's gravitational constant) exceeds the observed outflow line-of-sight speed of the gas, but Keplerian rotation is not encoded in absorption, and projection effects are also important. The small width of the absorption lines relative to the escape velocity may also indicate a low volume filling factor, consistent with a clumpy outflow or shell.

The existing observations show a general trend towards higher outflow speeds with time. Corresponding changes in ionization and column density are more modest, and not clearly linked to outflow speed[24]. However, some recent work has predicted higher outflow speeds in an initial super-Eddington disk regime, and lower outflow speeds in a subsequent thin disk regime[9,12]. An observation in an earlier, more highly super-Eddington phase might have observed broader lines and higher outflow speeds; future observations of new TDEs can test this.

Figure 3 shows the time evolution of the blackbody temperature measured in Swift/XRT monitoring observations. The temperature is remarkably constant, especially in contrast to the optical/ultraviolet decline shown in Fig. 1. Observations of steady blackbody temperatures, despite decaying multi-wavelength light curves in some TDEs[6,25], have recently been explained through winds[14]. Evidence of winds in our data supports this picture.

The low gas velocities may also be consistent with disrupted stellar gas on an elliptical orbit in a nascent disk, near the apocentre. This picture naturally gives a low filling factor, resulting in a small total mass in absorbing gas (see Methods). Recent numerical simulations predict that a fraction of the disrupted material in a TDE will circularize slowly[13], and that flows will be filamentary[26], while stellar gas that is more tightly bound can form an inner, Eddington-limited or super-Eddington disk more quickly.

The highly ionized, blueshifted gas discovered in our high-resolution X-ray spectra of ASASSN-14li confirms both fundamental and very recent theoretical predictions for the structure and evolution of TDEs. By pairing high-resolution X-ray spectroscopy with an ever-increasing number of TDE detections, it will become possible to test models of accretion disk formation and evolution, and to explore strong-field gravitation around massive black holes[27].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Bade, N., Komossa, S. & Dahlem, M. Detection of an extremely soft X-ray outburst in the HII-like nucleus of NGC 5905. *Astron. Astrophys.* **309,** L35–L38 (1996).
2. Komossa, S. & Greiner, J. Discovery of a giant and luminous X-ray outburst from the optically inactive galaxy pair RX J1242.6–1119. *Astron. Astrophys.* **349,** L45–L48 (1999).
3. Esquej, P. et al. Candidate tidal disruption events from the XMM-Newton slew survey. *Astron. Astrophys.* **462,** L49–L52 (2007).
4. Cappelluti, N. et al. A candidate tidal disruption event in the Galaxy cluster Abell 3571. *Astron. Astrophys.* **495,** L9–L12 (2009).
5. Gezari, S. et al. UV/optical detections of candidate tidal disruption events by GALEX and CFHTLS. *Astrophys. J.* **676,** 944–969 (2008).
6. Gezari, S. et al. An ultraviolet-optical flare from the tidal disruption of a helium-rich stellar core. *Nature* **485,** 217–220 (2012).
7. Guillochon, J., Manukian, H. & Ramirez-Ruiz, E. PS1–10jh: the disruption of a main-sequence star of near-solar composition. *Astrophys. J.* **783,** 23 (2014).
8. Rees, M. J. Tidal disruption of stars by black holes of 10^6–10^8 solar masses in nearby galaxies. *Nature* **333,** 523–528 (1988).
9. Strubbe, L. E. & Quataert, E. Optical flares from the tidal disruption of stars by massive black holes. *Mon. Not. R. Astron. Soc.* **400,** 2070–2084 (2009).
10. Lodato, G., King, A. R. & Pringle, J. E. Stellar disruption by a supermassive black hole: is the light curve really proportional to $t^{-5/3}$? *Mon. Not. R. Astron. Soc.* **392,** 332–340 (2009).
11. Lodato, G. & Rossi, E. M. Multiband light curves of tidal disruption events. *Mon. Not. R. Astron. Soc.* **410,** 359–367 (2011).
12. Strubbe, L. E. & Quataert, E. Spectroscopic signatures of the tidal disruption of stars by massive black holes. *Mon. Not. R. Astron. Soc.* **415,** 168–180 (2011).
13. Shiokawa, H., Krolik, J. H., Cheng, R. M., Piran, T. & Noble, S. C. General relativistic hydrodynamic simulation of accretion flow from a stellar tidal disruption. *Astrophys. J.* **804,** 85 (2015).
14. Miller, M. C. Disk winds as an explanation for slowly evolving temperatures in tidal disruption events. *Astrophys. J.* **805,** 83 (2015).
15. Jose, J. et al. ASAS-SN discovery of an unusual nuclear transient in PGC 043234. *Astron. Telegr.* **6777,** 1 (2014).
16. Gehrels, N. et al. The Swift Gamma-Ray Burst Mission. *Astrophys. J.* **611,** 1005–1020 (2004).
17. Burrows, D. N. et al. The Swift X-ray telescope. *Space Sci. Rev.* **120,** 165–195 (2005).

18. Voges, W. *et al.* The ROSAT All-Sky Survey bright source catalogue. *Astron. Astrophys.* **349,** 389–405 (1999).
19. Phinney, E. S. in *The Center of the Galaxy* (ed. Morris, M.) *IAU Symp.,* **136,** 543–553 (Kluwer Academic, 1989).
20. Kaastra, J. S., Mewe, R. & Nieuwenhuijzen, H. in *UV and X-ray Spectroscopy of Astrophysical and Laboratory Plasmas* (eds Yamashita, K. & Watanabe, T.) 411–414 (Universal Academy Press, Tokyo, 1996).
21. Loeb, A. & Ulmer, A. Optical appearance of the debris of a star disrupted by a massive black hole. *Astrophys. J.* **489,** 573–578 (1997).
22. Piran, T., Svirski, G., Krolik, J., Cheng, R. M. & Shiokawa, H. Disk formation versus disk accretion—what powers tidal disruption events? *Astrophys. J.* **806,** 164 (2015).
23. Shaviv, N. J. The theory of steady-state super-Eddington winds and its application to novae. *Mon. Not. R. Astron. Soc.* **326,** 126–146 (2001).
24. Ramírez, J. M. Kinematics from spectral lines for AGN outflows based on time-independent radiation-driven wind theory. *Rev. Mex. Astron. Astrofis.* **47,** 385–399 (2011).
25. Holoien, T. W.-S. *et al.* ASASSN-14ae: a tidal disruption event at 200 Mpc. *Mon. Not. R. Astron. Soc.* **445,** 3263–3277 (2014).
26. Guillochon, J. & Ramirez-Ruiz, E. A dark year for tidal disruption events. *Astrophys. J.* **809,** 166 (2015).
27. Stone, N. & Loeb, A. Observing Lense-Thirring precession in tidal disruption flares. *Phys. Rev. Lett.* **108,** 061302 (2012).
28. Kaastra, J. S., Mewe, R. & Raassen, T. New results on X-ray models and atomic data. *Highlights Astron.* **13,** 648–650 (2005).

METHODS

Estimates of prior black-hole luminosity. Using the ROSAT All-Sky Survey[18], the region around the host galaxy, PGC 043234, was searched for point sources. No sources were found. Points in the vicinity of the host galaxy were examined to derive a background count rate of 0.002 counts s^{-1}. Assuming the Milky Way column density $N_{\rm H,MW}$ along this line of sight, and taking a typical Seyfert X-ray spectral index of $\Gamma = 1.7$, this count rate translates into $L \approx 4.8 \times 10^{40}$ erg s^{-1}. This limit is orders of magnitude below a Seyfert or quasar luminosity.

Optical/ultraviolet monitoring observations and data reduction. Swift[16] monitors transient and variable sources via co-aligned X-ray (XRT, 0.3–10 keV) and ultraviolet–optical (UVOT, 170–650 nm) telescopes. High-cadence monitoring of ASASSN-14li with UVOT has continued in six bands: V, B, U, UVW1, UVM2, and UVW2 (central wavelength $\lambda_{\rm c} = 550$ nm, 440 nm, 350 nm, 260 nm, 220 nm and 190 nm).

All observations were processed using the latest HEASOFT (http://heasarc.gsfc.nasa.gov/docs/software/lheasoft/) suite and calibrations. Individual optical/ultraviolet exposures were astrometrically corrected and sub-exposures in each filter were summed. Source fluxes were then extracted from an aperture of 3″ radius, and background fluxes were extracted from a source-free region to the east of ASASSN-14li owing to the presence of a (blue) star lying 10 arcsec to the South, using UVOTMAGHIST, a routine within HEASOFT.

To estimate the host contamination, we have measured the host flux in 3″ aperture (matched to the aperture used for the UVOT photometry) in pre-outburst Sloan Digital Sky Survey (SDSS[29]), 2 Micron All-Sky Survey (2MASS[30]), and GALEX[31] images. We took extra care to deblend the GALEX data, where the large point spread function (PSF) resulted in contamination from the star about 10″ to the South. We estimated the uncertainty in each host flux by varying the inclusion aperture from 2″ to 4″.

We then fitted the host photometry to synthetic galaxy templates using the Fitting and Assessment of Synthetic Templates (FAST[32]) code. We employed stellar templates from the[33] catalogue, and allowed the star-formation history, extinction law, and initial mass function to vary over the full range of parameters allowed by the software. All best-fit models had stellar masses of about $10^{9.2} M_\odot$, low ongoing star-formation rates (at most about $10^{-1.5} M_\odot$ yr^{-1}), and modest line-of-sight extinction ($A_{\rm V} \lesssim 0.4$ mag).

We integrated the resulting galaxy template spectra over each UVOT filter bandpass to estimate the host count rate. For the uncertainty in this value, we adopt either the root-mean-square spread of the resulting galaxy template models, or 10% of the inferred count rate, whichever value was larger. We then subtracted these values from our measured (coincidence-loss corrected) photometry of the host plus transient, to isolate the component that is due to TDE. For reference, our inferred count rates for each UVOT filter are 5.7 ± 0.6 s^{-1} for V, 9.4 ± 0.9 s^{-1} for B, 4.0 ± 0.4 s^{-1} for U, 0.83 ± 0.08 s^{-1} for UVW1, 0.29 ± 0.03 s^{-1} for UVM2, and 0.49 ± 0.05 s^{-1} for UVW2. Figure 1 shows the host-subtracted optical and ultraviolet light curves ASASSN-14li.

Fits to the UVOT/UVM2 light curve. The UVM2 filter provides the most robust trace of the mass accretion rate in a TDE like ASASSN-14li; it has negligible transmission at optical wavelengths[34,35]. Fits to the UVM2 light curve with a power law of the form $f(t) = f_0 \times (t + t_0)^{\alpha}$ with a fixed index of $\alpha = -5/3$ imply a disruption date of $t_0 = 56{,}980 \pm 3$ (MJD). This model achieves a fair characterization of the data; high fluxes between days 80 and 100 (in the units of Fig. 1) result in a poor statistical fit ($\chi^2/\nu = 1.7$, where $\nu = 54$ degrees of freedom). If the light curve is fitted with a variable index, a value of -2.6 ± 0.3 is measured (90% confidence). This model achieves an improved fit ($\chi^2/\nu = 1.4$, for $\nu = 53$ degrees of freedom), but it does not tightly constrain the disruption date, placing t_0 in the MJD 56,855–56,920 range. That disruption window is adjacent to an interval wherein the ASASSN monitoring did not detect the source[15], making it less plausible than the fit with $\alpha = -5/3$.

The optical bands appear to have a shallower decay curve than the ultraviolet bands. Recent theory[11] predicts that optical light produced via thermal disk emission should show a decay consistent with $t^{-5/12}$; this might also be due to reprocessing[7]. The V-band data are consistent with this prediction, though the data are of modest quality and a broad range of decays are permitted.

X-ray monitoring observations and data reduction. The Swift XRT[17] is a charge-coupled device. In such cameras, photon pile-up occurs when two or more photons land within a single detection box during a single frame time. This causes flux distortions and spectral distortions to bright sources. Such distortions are effectively avoided by extracting events from an annular region, rather than from a circle at the centre of the telescope PSF. We therefore extracted source spectra from annuli with an inner radius of 12 arcsec (5 pixels), and an outer radius of 50 arcsec. Background flux was measured in an annular region extending from 140 arcsec to 210 arcsec.

Standard redistribution matrices were used; an ancillary response file was created with the *xrtmkarf* tool (a routine within HEASOFT) using a vignetting corrected exposure map. The source spectra were rebinned to have 20 counts per bin with *grppha*. In all spectral fits, we adopted a lower spectral bound of 0.3 keV (36 Å). The upper bound on spectral fits varied depending on the boundary of the last bin with at least 20 counts; this was generally around 1 keV (12 Å). The XRT spectra were fitted with a model consisting of absorption in the Milky Way of a blackbody emitted at the redshift of the TDE, that is, *pha(zashift(bbodyrad))*, where $N_{\rm H} \equiv 4 \times 10^{20}$ cm^{-2} and $z \equiv 0.0206$. The evolution of the best-fit temperature of this blackbody component is displayed in Fig. 3.

The blackbody temperature values measured from the Swift XRT are slightly higher ($kT \approx 7$–10 eV) than those measured with XMM-Newton and Chandra. If an outflow component with fiducial parameters is included in the spectral model anyway, the XRT temperatures are then in complete agreement with those measured using XMM-Newton and Chandra.

Estimates of the black-hole mass. Luminosity values inferred for the band over which the high-resolution spectra are actually fitted, and for a broader band, are listed in Table 1. Taking the broader values as a proxy for a true bolometric fit, the highest implied soft X-ray luminosity is measured in the last XMM-Newton monitoring observation, giving $L \approx 3.2 \times 10^{44}$ erg s^{-1}. The Eddington luminosity for standard hydrogen-rich accretion is $L_{\rm Edd} = 1.3 \times 10^{38}$ erg s^{-1} $(M/M_\odot)$. This implies a black-hole mass of $M \approx 2.5 \times 10^6 M_\odot$.

Blackbody continua imply size scales, and, if we assume that optically thick blackbody emission can only originate at radii larger than the innermost stable circular orbit (ISCO), also masses. For a non-spinning Schwarzschild black hole, $r_{\rm ISCO} = 6GM/c^2$. The blackbody emission measured in fits to the time-averaged XMM-Newton 'long stare' gives an emitting area of 3.7×10^{25} cm^2; implying $r = 1.7 \times 10^{12}$ cm for a spherical geometry. The actual geometry may be more disk-like, but the inner flow may be a thick disk that is better represented by a spherical geometry. If the black hole powering ASASSN-14li is not spinning, this size implies a black-hole mass of $M \approx 1.9 \times 10^6 M_\odot$.

We also estimated the mass of the black hole at the heart of ASASSN-14li by fitting the host-subtracted light curves (see Fig. 1) using the Monte Carlo software TDEFit[7]. This software assumes that emission is produced within an elliptical accretion disk where the mass accretion rate follows the fallback rate[36] onto the black hole with a viscous delay[26]. This emission is then partly reprocessed into the ultraviolet/optical part of the spectrum by an optically thick layer[21]. Super-Eddington accretion is treated by presuming that a fitted fraction of the Eddington excess is converted into light that is reprocessed by the same optically thick layer. This excess can be produced either with an unbound wind[9,37], or with the energy deposited by shocks in the circularization process[13,22].

The software performs a maximum-likelihood analysis to determine the combinations of parameters that reproduce the observed light curves. We utilize the ASASSN, UVOT and XRT data in our light-curve fitting; the most likely models produce good fits to all bands simultaneously. Within the context of this TDE model, a black-hole mass of $(0.4$–$1.2) \times 10^6 M_\odot$ (1σ) is derived.

Spectroscopic observations, data reduction and analysis. Table 1 lists the observation identification number, start time, and duration of all of the XMM-Newton and Chandra observations considered in our work.

The XMM-Newton data were reduced using the standard Science Analysis System (SAS version 13.5.0) tools and the latest calibration files. The *rgsproc* routine was used to generate spectral files from the source, background spectral files, and instrument response files. The spectra from the RGS1 and RGS2 units were fitted jointly. Prior to fitting models, all XMM-Newton spectra were binned by a factor of five for clarity and sensitivity.

The Chandra data were reduced using the standard Chandra Interactive Analysis of Observations (CIAO version 4.7) suite, and the latest associated calibration files. Instrument response files were constructed using the *fullgarf* and *mkgrmf* routines. The first-order spectra from each observation were combined using the tool *add grating orders*, and spectra from each observation were then added using *add grating spectra*.

The spectra were analysed using the SPEX suite version 2.06 (ref. 20). The fitting procedure minimized a χ^2 statistic. The spectra are most sensitive in the 18–35 Å band, and all fits were restricted to this range. Within SPEX, absorption from the interstellar medium in the Milky Way was modelled using the model 'hot'; a separate 'hot' component was included to allow for interstellar medium (ISM) absorption within PGC 043234 at its known redshift (using the *reds* component in SPEX). The photoionized outflow was modelled using the *pion* component within the SPEX suite.

pion[20] includes numerous lines from intermediate charge states that are lacking in similar astrophysics packages. The fits explored in this analysis varied the gas column density ($N_{\rm H,TDE}$), the gas ionization parameter (ξ, where $\xi = L/nr^2$, and L is luminosity, n is the hydrogen number density and r is the distance between the

ionizing source and absorbing gas), the root-mean-square velocity of the gas (v_{rms}), and the bulk shift of the gas relative to the source, in the source frame (v_{shift}). Spectra from segments within the 'long stare' made with XMM-Newton were made by using the SAS tool *tabgtigen* to create *good time interval* files to isolate periods within the light curves of the RGS data.

The Chandra/LETG spectra were dispersed onto the HRC, which has a relatively high instrumental background. Fitting the spectra only in the 18–35 Å band served to limit the contributions of the background. Nevertheless, the Chandra spectra are less sensitive than the best XMM-Newton spectra of ASASSN-14li (see Fig. 2). Prior to fitting, spectra from the two exposures were added and then binned by a factor of three.

Figure 2 includes plots of the $\Delta\chi^2$ goodness-of-fit statistic as a function of wavelength, before and after including *pion* to model the ionized absorption. There is weak evidence of emission lines in the spectra, perhaps with a P Cygni profile (see below). The best-fit models for the high-resolution spectra predict one absorption line at 34.5 Å (H-like C VI) that is not observed; small variations to abundances could resolve this disparity.

Blueshifts as small as $200\,\mathrm{km\,s^{-1}}$ are measured in the XMM-Newton/RGS using the *pion* model. According to the XMM-Newton User's Handbook, available through the mission website, http://xmm.esac.esa.int/external/xmm_user_support/documentation/index.shtml, the absolute accuracy of the first-order wavelength scale is 6 mÅ. At 18 Å this corresponds to a velocity of $100\,\mathrm{km\,s^{-1}}$; at 35 Å, this corresponds to a velocity of $51\,\mathrm{km\,s^{-1}}$. The model predicts numerous lines across the 18–35 Å band that are clearly detected; especially with this leverage, the small shifts we have measured with XMM-Newton are robust. In particular, the difference in blueshift between the low- and high-flux phases of the 'long stare', $-360 \pm 50\,\mathrm{km\,s^{-1}}$ versus $-130^{+70}_{-50}\,\mathrm{km\,s^{-1}}$, is greater than the absolute calibration uncertainties. Differences observed in the outflow velocities between XMM-Newton observations are as large, or larger, and also robust.

The lower sensitivity of the Chandra spectra is evident in the relatively poor constraints achieved on the column density of the ionized X-ray outflow $N_{H,TDE}$ (see Table 1). Similarly, the relatively high outflow velocity measured in the Chandra spectra should be viewed with a degree of caution. The outflow velocity changes from about $500\,\mathrm{km\,s^{-1}}$ to just $-130 \pm 130\,\mathrm{km\,s^{-1}}$, for instance, when the binning factor is increased from three to five. We have found no reports in the literature of a systematic wavelength offset between contemporary high-resolution spectra obtained with XMM-Newton and Chandra.

The small number of high-resolution spectra complicates efforts to discern trends. The velocity width of the absorbing gas is fairly constant over time, but there is a general trend towards higher blue-shifts. There is no clear trend in column density or ionization parameter with time.

Diffuse gas mass, outflow rates and filling factors. There is no a priori constraint on the density of the absorbing gas. Taking the maximum radius implied by variability within XMM-Newton 'long stare', $r \leq 3 \times 10^{15}\,\mathrm{cm}$, and manipulating the ionization parameter equation ($\xi = Ln^{-1}r^{-2}$, where L is the luminosity, n is the number density and r is the absorbing radius), we can derive an estimate of the density: $n \approx 2 \times 10^9\,\mathrm{cm^{-3}}$. Even assuming a uniformly filled sphere out to a radius of $r = 3 \times 10^{15}\,\mathrm{cm}$, a total mass of $M \approx 4 \times 10^{32}\,\mathrm{g}$ is implied, or approximately $0.2 M_\odot$.

The true gas mass within r is likely to be orders of magnitude lower, owing to clumping and a very low volume filling factor. Using the measured value of $N_{H,TDE}$ and assuming $n \approx 2 \times 10^9\,\mathrm{cm^{-3}}$, $N_{H,TDE} = n\Delta r$ gives a value of $\Delta r \approx 6.5 \times 10^{12}\,\mathrm{cm}$. The filling factor can be estimated using $\Delta r/r \approx 0.002$. The total mass enclosed out to a distance r is then reduced accordingly, down to $4 \times 10^{-4} M_\odot$, assuming a uniform density within r. This is a small value, plausible either for a clumpy wind or gas within a filament executing an elliptical orbit.

Formally, the mass outflow rate in ASASSN-14li can be adapted from the case where the density is known, and written as:

$$\dot{M}_{out} = \mu m_p \Omega L v C_v \xi^{-1}$$

where μ is the mean atomic weight ($\mu = 1.23$ is typical), m_p is the mass of the proton, Ω is the covering factor ($0 \leq \Omega \leq 4\pi$), L is the ionizing luminosity, v is the outflow velocity, C_v is the line-of-sight global (volume) filling factor and ξ is the ionization parameter. Using the values obtained in fits to the XMM-Newton 'long stare' (see Table 1), for instance, $\dot{M}_{out} \approx 7.9 \times 10^{23} \Omega C_v\,\mathrm{g\,s^{-1}}$. Taking the value of C_v derived above, an outflow rate of $\dot{M}_{out} \approx 1.5 \times 10^{21} \Omega\,\mathrm{g\,s^{-1}}$ results. The kinetic power in the outflow is given by $L_{kin} = 0.5\dot{M}v^2$; using the same values assumed to estimate the mass outflow rate, $L_{kin} \approx 3.3 \times 10^{35}\,\mathrm{erg\,s^{-1}}$.

Emission from the diffuse outflow. We synthesized a plausible wind emission spectrum by coupling the *pion* and *hyd* models within SPEX. The *hyd* code enables spectra to be constructed based on the output of hydrodynamical simulations. As inputs, the *hyd* code requires the electron temperature and ion concentrations for a gas; these were taken from our fits with *pion*. We included the resulting emission component in experimental fits to the XMM-Newton 'long stare'. The best-fit model gives an emission measure of $(1.0 \pm 0.3) \times 10^{64}\,\mathrm{cm^{-3}}$, a redshift (relative to the host) of $270^{+350}_{-150}\,\mathrm{km\,s^{-1}}$, and an ionization parameter of $\log\xi = 4.3 \pm 0.1$.

According to an F-test, the emission component is only required at the 3σ level; however, it has some compelling properties. Combined with the blueshifted absorption spectrum, the redshifted emission gives P Cygni profiles. For the gas density of $n \approx 2 \times 10^9\,\mathrm{cm^{-3}}$ derived previously, the emission measure gives a radius of about $10^{15}\,\mathrm{cm}$, comparable to the size scale inferred from absorption variability.

The strongest lines predicted by the emission model include He-like O VII, and H-like charge states of C, N and O. This model does not account for other emission line-like features in the spectra, which are more likely to be artefacts from spectral binning, or calibration or modelling errors. Emission features in the O K-edge region may be real, but caution is warranted. Other features are more easily discounted given that they differ between the RGS1 and RGS2 spectra.

Code availability. All of the data reduction and spectroscopic fitting routines and packages used in this work are publicly available. The light-curve modelling package, TDEFit[7], is proprietary at this time owing to ongoing code development; a public release is planned within the coming year.

29. Ahn, C. P. *et al.* The Tenth Data Release of the Sloan Digital Sky Survey: first spectroscopic data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment. *Astrophys. J. Suppl. Ser.* **211**, 17 (2014).
30. Skrutskie, M. F. *et al.* The Two Micron All Sky Survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
31. Martin, D. C. *et al.* The Galaxy Evolution Explorer: a space ultraviolet survey mission. *Astrophys. J.* **619**, L1–L6 (2005).
32. Kriek, M. *et al.* An ultra-deep near-infrared spectrum of a compact quiescent galaxy at z = 2.2. *Astrophys. J.* **700**, 221–231 (2009).
33. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
34. Poole, T. S. *et al.* Photometric calibration of the Swift ultraviolet/optical telescope. *Mon. Not. R. Astron. Soc.* **383**, 627–645 (2008).
35. Breeveld, A. A. *et al.* An updated ultraviolet calibration for the Swift/UVOT. *AIP Conf. Ser.* (eds McEnery, J. E., Racusin, J. L. & Gehrels, N.), **1358**, 373–376 (American Institute of Physics, 2011).
36. Guillochon, J. & Ramirez-Ruiz, E. Hydrodynamical simulations to determine the feeding rate of black holes by the tidal disruption of stars: the importance of the impact parameter and stellar structure. *Astrophys. J.* **767**, 25 (2013).
37. Vinkó, J. *et al.* A luminous, fast rising UV-transient discovered by ROTSE: a tidal disruption event? *Astrophys. J.* **798**, 12 (2015).

LETTER

A disintegrating minor planet transiting a white dwarf

Andrew Vanderburg[1], John Asher Johnson[1], Saul Rappaport[2], Allyson Bieryla[1], Jonathan Irwin[1], John Arban Lewis[1], David Kipping[1,3], Warren R. Brown[1], Patrick Dufour[4], David R. Ciardi[5], Ruth Angus[1,6], Laura Schaefer[1], David W. Latham[1], David Charbonneau[1], Charles Beichman[5], Jason Eastman[1], Nate McCrady[7], Robert A. Wittenmyer[8] & Jason T. Wright[9]

Most stars become white dwarfs after they have exhausted their nuclear fuel (the Sun will be one such). Between one-quarter and one-half of white dwarfs have elements heavier than helium in their atmospheres[1,2], even though these elements ought to sink rapidly into the stellar interiors (unless they are occasionally replenished)[3–5]. The abundance ratios of heavy elements in the atmospheres of white dwarfs are similar to the ratios in rocky bodies in the Solar System[6,7]. This fact, together with the existence of warm, dusty debris disks[8–13] surrounding about four per cent of white dwarfs[14–16], suggests that rocky debris from the planetary systems of white-dwarf progenitors occasionally pollutes the atmospheres of the stars[17]. The total accreted mass of this debris is sometimes comparable to the mass of large asteroids in the Solar System[1]. However, rocky, disintegrating bodies around a white dwarf have not yet been observed. Here we report observations of a white dwarf—WD 1145+017—being transited by at least one, and probably several, disintegrating planetesimals, with periods ranging from 4.5 hours to 4.9 hours. The strongest transit signals occur every 4.5 hours and exhibit varying depths (blocking up to 40 per cent of the star's brightness) and asymmetric profiles, indicative of a small object with a cometary tail of dusty effluent material. The star has a dusty debris disk, and the star's spectrum shows prominent lines from heavy elements such as magnesium, aluminium, silicon, calcium, iron, and nickel. This system provides further evidence that the pollution of white dwarfs by heavy elements might originate from disrupted rocky bodies such as asteroids and minor planets.

WD 1145+017 (also designated EPIC 201563164) is a helium-envelope white dwarf (Supplementary Table 1) that was observed by NASA's Kepler space telescope during the first campaign of its two-wheeled mission—a mission referred to hereafter as K2. After processing K2 data taken from WD 1145+017 to produce a light curve and correcting for instrumental systematics[18], we identified a transit-like signal with a period of 4.5 h by using a box-fitting least-squares search algorithm[19]. Using a Fourier analysis on the systematic-corrected K2 data, we identified five other weaker, but statistically significant, periodicities in the data, all with periods between 4.5 h and 5 h (Fig. 1 and Supplementary Table 2). We examined the dominant periodicity and found that the depth and shape of the transits varied substantially over the 80 days of K2 observations (Fig. 2).

We initiated follow-up, ground-based photometry to achieve better time resolution of the transits seen in the K2 data (Supplementary Fig. 1). We observed WD 1145+017 frequently over the course of about a month with the 1.2-metre telescope at the Fred L. Whipple Observatory (FLWO) on Mount Hopkins, Arizona; with one of the 0.7-metre MINiature Exoplanet Radial Velocity Array (MINERVA)

telescopes, also at FLWO; and with four of the eight 0.4-metre telescopes that compose the MEarth–South Array at the Cerro Tololo Inter-American Observatory in Chile. Most of these data showed no interesting or noteworthy signals, but on two nights we observed deep (with up to 40% of the star's brightness blocked), short-duration (5-min), asymmetric transits separated by the dominant 4.5-h period identified in the K2 data (Fig. 3). In particular, using the 1.2-m FLWO telescope in the V-band (green visible light), we detected two transits separated by 4.5 h on the night of 11 April 2015; furthermore, using four of the eight MEarth–South array telescopes (all in near-infrared light, using a 715-nm long-pass filter), we detected two transits separated by the same 4.5-h period on the night of 17 April 2015. The transits did not occur at the times predicted from the K2 data, and the two transits detected on 11 April happened nearly 180 degrees out of phase from the two transits detected on 17 April. Observations with MEarth–South in near-infrared light and with MINERVA in white visible light the next night (18 April) showed only a possible transit event, of 10%–15% depth, in phase with the previous night's events. The 5-min duration of the transits is longer than the roughly 1-min duration we would expect for a solid body transiting the white dwarf.

Nonetheless, we confirmed that these events are indeed transits by a low-mass object in orbit around the white dwarf. The depth and morphology of the transits that we see in the ground-based data cannot be explained by stellar pulsations, and archival and adaptive optics imaging place strong constraints on scenarios involving a binary star in the background, whose eclipses might mimic transits of the white dwarf (Supplementary Fig. 2). We also obtained spectroscopic observations with the MMT Blue Channel spectrograph; we used these observations to place limits on radial-velocity variations that would indicate stellar companions. The radial-velocity measurements exclude companions larger than ten Jupiter masses at the 95% confidence level.

The spectra also reveal that the atmosphere of the white dwarf contains magnesium, aluminium, silicon, calcium, iron, and nickel (Supplementary Fig. 3). These elements, which are heavier than helium, have settling times that are much shorter than the cooling age of the white dwarf, indicating that they have been deposited in the white dwarf's envelope in the past million years[5]—much more recently than it formed, about 175 ± 75 million years ago. Archival photometry for this system is well fitted by a 15,900-K, metal-rich white-dwarf model spectrum, and we find evidence for excess infrared emission consistent with a warm (1,150 K) dusty debris disk (Supplementary Fig. 4).

We interpret these observations as evidence that at least one, and probably six or more, disintegrating planetesimals are transiting this white dwarf. Disintegrating planets have been observed transiting main-sequence stars[20–22], and show asymmetric transit profiles and

[1]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. [2]Department of Physics, and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Astronomy, Columbia University, New York, New York 10027, USA. [4]Institut de Recherche sur les Exoplanètes, Départment de Physique, Université de Montréal, Montréal, Quebec H3C 3J7, Canada. [5]NASA Exoplanet Science Institute, California Institute of Technology, Pasadena, California 91125, USA. [6]Department of Physics, University of Oxford, Oxford OX1 3RH, UK. [7]Department of Physics and Astronomy, University of Montana, Missoula, Montana 59812, USA. [8]School of Physics and Australian Centre for Astrobiology, University of New South Wales, Sydney, New South Wales 2052, Australia. [9]Department of Astronomy and Astrophysics and Center for Exoplanets and Habitable Worlds, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.
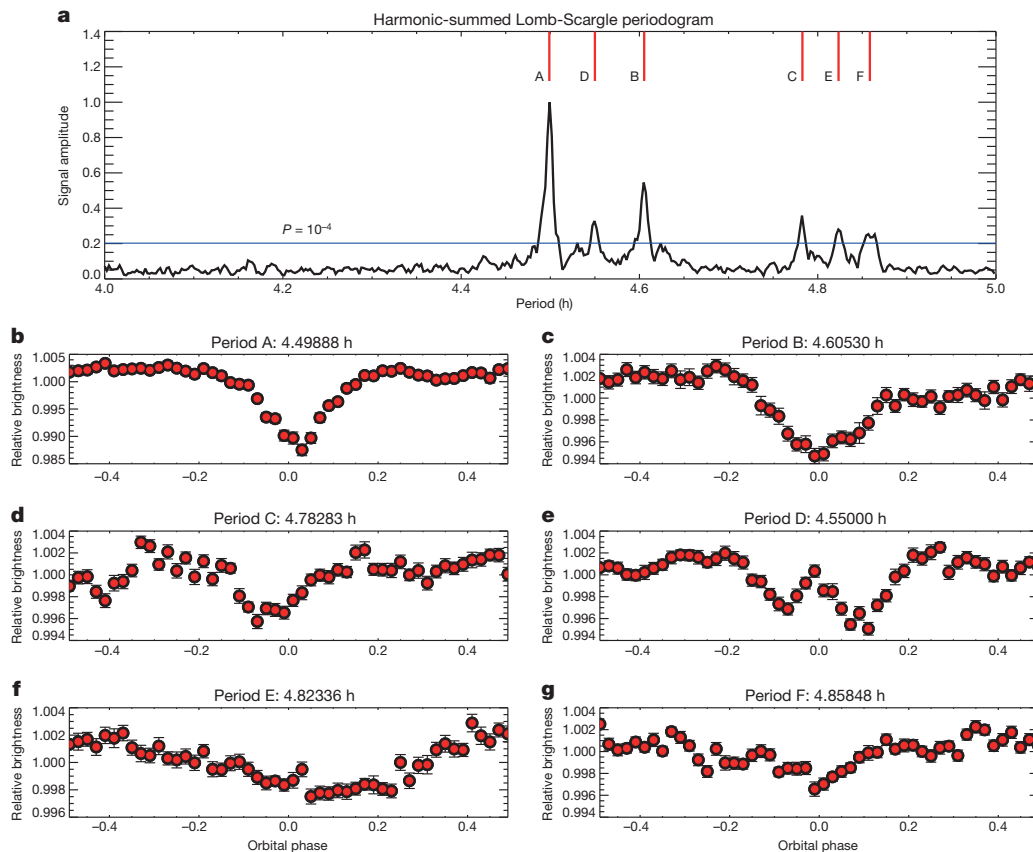
Figure 1 | Six notable periodicities found in the K2 data. a, Harmonic-summed Lomb–Scargle periodogram of the K2 data. We label the signals A–F in order of amplitude. **b–g,** K2 light curves folded on the six statistically significant ($P < 10^{-4}$) peaks and binned in phase. When plotting each fold, we sequentially removed stronger signals by dividing the data set by the binned, phase-folded light curves of the stronger signals. Note the differences in scales on the y-axes. Error bars are the standard errors of the mean within each bin. Brightness is shown relative to the median brightness measurement of WD 1145+017.

variable transit depths—behaviours similar to those seen here. These previously detected disintegrating planets are believed to be heated by the host star and to be losing mass through Parker-type thermal winds, in which the molecules condense into the obscuring dust observed to be occulting the star[23]. The solid bodies themselves are too small to detect, so the transits are dominated by the much larger dust clouds trailing the planets. The density of the dust cloud is presumed to be highly variable, which gives rise to the variable transit depths; in addition, a comet-like structure for the dust tails would explain the asymmetric transit shapes[20–22]. In the case of WD 1145+017, we have identified six stable periodicities in the K2 light curve that could be explained by occultations of the central star by dust clouds. We propose that each of these periodicities could be related to a different planetesimal (or to multiple fragments of one minor planet) that is orbiting the white dwarf near the tidal radius for rocky bodies. Each planetesimal would sporadically launch winds of metal gases, which are probably streaming freely from the planetesimal and which condense into dust clouds that periodically block the light of the white dwarf. A trailing dust cloud would explain the variable transit depths, the asymmetric transit profiles, and the longer-than-expected transit durations that we see in the light curves of WD 1145+017 (Supplementary Fig. 5).

We have simulated the dynamics of six planetesimals in circular orbits with periods of between 4.5 h and 4.9 h, and find that such a configuration is stable for at least 10^6 orbits, provided that their masses are smaller than or comparable to that of the dwarf planet Ceres ($1.6 \times 10^{-4} M_\oplus$, where $M_\oplus$ is the mass of the Earth), or possibly that of Haumea ($6.7 \times 10^{-4} M_\oplus$). These six planetesimals must be rocky (because gaseous bodies would overflow their Roche lobes, the region within which gaseous material can be stably retained by gravity), and

must have densities greater than about $2 \, \mathrm{g \, cm^{-3}}$ in order not to be tidally disrupted in such short-period orbits[24,25]. We also simulated the dynamics of two planetesimals in 1:1 mean motion orbital resonances (for example, in horseshoe orbits), and find that two different planetesimals in such orbits outbursting at different times could plausibly explain the difference in orbital phases that we see between the K2 light curve, the 11 April events, and the 17 April events.

We estimate that a rate of mass loss of roughly $8 \times 10^9 \, \mathrm{g \, s^{-1}}$ is necessary to explain the transits that we see. Various refractory materials (including iron, fayalite, albite, and orthoclase) heated by the white dwarf could plausibly sublimate from a planetesimal roughly the size of Ceres at this rate, despite the white dwarf's relatively low luminosity (Supplementary Fig. 6). These metal vapours would be lost quickly via free-streaming winds or by Jeans escape (a classical thermal escape mechanism), because the planetesimal escape velocity is comparable to the metal vapour's thermal speed. We simulated a dust cloud condensed from the escaped metal vapour in orbit[20–22], and found that the radiation environment in which these planetesimals are situated can give rise to dust tails like those we infer from the ground-based transit observations (Supplementary Fig. 7). Collisions with disk debris[26] could also plausibly cause mass from the planetesimal to be lost into orbit.

A possible scenario for the formation of the disintegrating planetesimals involves minor planets that are left over from the progenitor stellar system, before the star evolved into a white dwarf[4,17]. In this scenario, mass loss from the host star disturbs the stability of the planetary system. This can lead to planets or smaller objects (such as asteroids or comets) being scattered inwards, into orbits with radii much smaller than the size of the progenitor star when it was an
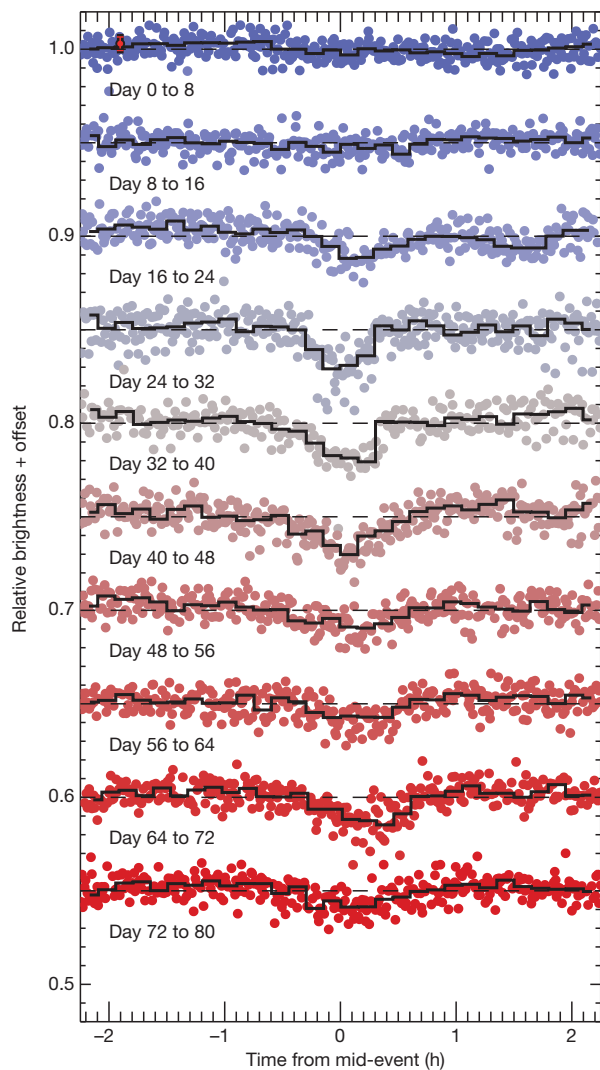
Figure 2 | Evolution of the K2 transit light curve over 80 days of observations. We show the K2 light curve broken into segments eight days in length and folded on the most notable, 4.5-h period. The individual data points (sampled with a 30-min integration time) are shown as dots, with different colours representing different time segments. The averaged light curve for each bin is shown as a solid black line. Each segment is vertically offset for clarity. We show the typical measurement uncertainty (standard deviation) with a red error bar on one data point in the upper left.

evolving giant. A challenge for this model is placing the planetesimals in close concentric orbits so near the star without being totally disrupted. Current models suggest that planetesimals can be scattered inwards on highly eccentric orbits, tidally disrupted into elliptical dust disks, and circularized by Poynting–Robertson drag[27]. However, bodies that could release enough dust to cause the transits of WD 1145+017 that we have detected are too large to be circularized in this way. Recent theoretical work[28] on smaller bodies has shown that outgassing material can quickly circularize orbits, but it is unclear how this process scales to the massive bodies inferred here.

Our interpretation of this system is still uncertain. In particular, it is difficult to explain the phase shifts observed between the transits detected by ground-based photometry and those detected by K2; further ground-based observations are necessary to understand this effect. Another possible model is that small rings[27] or debris clouds of disrupted planetary material in a disk occasionally cross in front of the star and block its light. Although this could explain the large phase shifts that we see between the FLWO and MEarth transits, it is difficult to explain the highly stable periods ($\Delta P/P < 10^{-4}$) seen in the K2 data
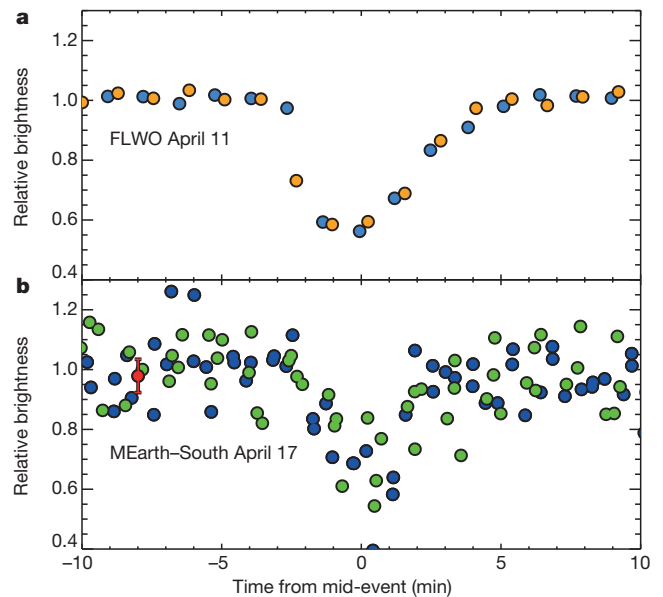
Figure 3 | Transit light curves measured from two ground-based facilities. a, Two events observed at FLWO with a separation equal to the 4.5-h period labelled A in Fig. 1, detected by K2. The first event is blue, and the second is orange. **b**, Two events observed by MEarth–South separated by the 4.5-h A-period. The first event is blue, and the second is green. The typical MEarth–South measurement uncertainty (standard deviation) is shown as a red error bar on one data point. The FLWO error bars are smaller than the size of the symbols.

without massive orbiting bodies (Supplementary Fig. 8). Fortunately, the large transit depths make follow-up observations that could distinguish among these scenarios feasible both from the ground and from space. It might be possible to detect periodic infrared emission[29] from the orbiting planetesimals with the James Webb Space Telescope. Additional follow-up observations such as transit spectroscopy could constrain both scenarios by detecting the presence of molecules in the dust tails or the wavelength dependence of the dust scattering[30].

The evidence presented here—in particular, for the heavy-element pollution of the white dwarf WD 1145+017, for a warm dusty debris disk around this star, and for transits of disintegrating planetesimals—is consistent with a scenario, suggested over the past decade, in which the orbits of rocky bodies are occasionally perturbed and pass close enough to white dwarfs to become tidally disrupted, leading to the infall of debris onto the star's surface. Observations have shown that this scenario could be quite common among white dwarfs, with between 25% and 50% of white dwarfs showing evidence of heavy-element pollution. Our observations indicate that disintegrating planetesimals may be common as well (Supplementary Fig. 9). The transits of WD 1145+017 provide evidence of rocky, disintegrating bodies around a white dwarf, and support the planetesimal accretion model for the pollution of white dwarfs.

1. Zuckerman, B., Melis, C., Klein, B., Koester, D. & Jura, M. Ancient planetary systems are orbiting a large fraction of white dwarf stars. *Astrophys. J.* **722,** 725–736 (2010).
2. Koester, D., Gänsicke, B. T. & Farihi, J. The frequency of planetary debris around young white dwarfs. *Astron. Astrophys.* **566,** A34 (2014).
3. Paquette, C., Pelletier, C., Fontaine, G. & Michaud, G. Diffusion in white dwarfs—new results and comparative study. *Astrophys. J.* **61** (Suppl.), 197–217 (1986).
4. Aannestad, P. A., Kenyon, S. J., Hammond, G. L. & Sion, E. M. Cool metallic-line white dwarfs, radial velocities, and interstellar accretion. *Astron. J.* **105,** 1033–1044 (1993).
5. Koester, D. Accretion and diffusion in white dwarfs. New diffusion timescales and applications to GD 362 and G 29–38. *Astron. Astrophys.* **498,** 517–525 (2009).
6. Zuckerman, B., Koester, D., Melis, C., Hansen, B. M. & Jura, M. The chemical composition of an extrasolar minor planet. *Astrophys. J.* **671,** 872–877 (2007).

7. Farihi, J., Gänsicke, B. T. & Koester, D. Evidence for water in the rocky debris of a disrupted extrasolar minor planet. *Science* **342,** 218–220 (2013).
8. Zuckerman, B. & Becklin, E. E. Excess infrared radiation from a white dwarf—an orbiting brown dwarf? *Nature* **330,** 138–140 (1987).
9. Jura, M. A tidally disrupted asteroid around the white dwarf G29–38. *Astrophys. J.* **584,** L91–L94 (2003).
10. Becklin, E. E. *et al.* A dusty disk around GD 362, a white dwarf with a uniquely high photospheric metal abundance. *Astrophys. J.* **632,** L119–L122 (2005).
11. Kilic, M., von Hippel, T., Leggett, S. K. & Winget, D. E. Excess infrared radiation from the massive DAZ white dwarf GD 362: a debris disk? *Astrophys. J.* **632,** L115–L118 (2005).
12. Jura, M., Farihi, J., Zuckerman, B. & Becklin, E. E. Infrared emission from the dusty disk orbiting GD 362, an externally polluted white dwarf. *Astron. J.* **133,** 1927–1933 (2007).
13. Kilic, M., von Hippel, T., Leggett, S. K. & Winget, D. E. Debris disks around white dwarfs: the DAZ connection. *Astrophys. J.* **646,** 474–479 (2006).
14. Farihi, J., Jura, M. & Zuckerman, B. Infrared signatures of disrupted minor planets at white dwarfs. *Astrophys. J.* **694,** 805–819 (2009).
15. Barber, S. D. *et al.* The frequency of debris disks at white dwarfs. *Astrophys. J.* **760,** 26 (2012).
16. Rocchetto, M., Farihi, J., Gänsicke, B. T. & Bergfors, C. The frequency and infrared brightness of circumstellar discs at white dwarfs. *Mon. Not. R. Astron. Soc.* **449,** 574–587 (2015).
17. Debes, J. H. & Sigurdsson, S. Are there unstable planetary systems around white dwarfs? *Astrophys. J.* **572,** 556–565 (2002).
18. Vanderburg, A. & Johnson, J. A. A technique for extracting highly precise photometry for the two-wheeled Kepler mission. *Publ. Astron. Soc. Pacif.* **126,** 948–958 (2014).
19. Kovács, G., Zucker, S. & Mazeh, T. A box-fitting algorithm in the search for periodic transits. *Astron. Astrophys.* **391,** 369–377 (2002).
20. Rappaport, S. *et al.* Possible disintegrating short-period super-Mercury orbiting KIC 12557548. *Astrophys. J.* **752,** 1 (2012).
21. Rappaport, S. *et al.* KOI-2700b—a planet candidate with dusty effluents on a 22 hr orbit. *Astrophys. J.* **784,** 40 (2014).
22. Sanchis-Ojeda, R. *et al.* The K2-ESPRINT project I: discovery of the disintegrating rocky planet with a cometary head and tail EPIC 201637175b. Preprint at http://arxiv.org/abs/1504.04379 (2015).
23. Perez-Becker, D. & Chiang, E. Catastrophic evaporation of rocky planets. *Mon. Not. R. Astron. Soc.* **433,** 2294–2309 (2013).
24. Rappaport, S., Sanchis-Ojeda, R., Rogers, L. A., Levine, A. & Winn, J. N. The Roche limit for close-orbiting planets: minimum density, composition constraints, and application to the 4.2 hr planet KOI 1843.03. *Astrophys. J.* **773,** L15 (2013).
25. Davidsson, B. J. R. Tidal splitting and rotational breakup of solid spheres. *Icarus* **142,** 525–535 (1999).
26. Farihi, J., Zuckerman, B. & Becklin, E. E. Spitzer IRAC observations of white dwarfs. I. Warm dust at metal-rich degenerates. *Astrophys. J.* **674,** 431–446 (2008).
27. Veras, D., Leinhardt, Z. M., Bonsor, A. & Gänsicke, B. T. Formation of planetary debris discs around white dwarfs—I. Tidal disruption of an extremely eccentric asteroid. *Mon. Not. R. Astron. Soc.* **445,** 2244–2255 (2014).
28. Veras, D., Eggl, S. & Gänsicke, B. T. Sublimation-induced orbital perturbations of extrasolar active asteroids and comets: application to white dwarf systems. *Mon. Not. R. Astron. Soc.* **452,** 1945–1957 (2015).
29. Lin, H. W. & Loeb, A. Finding rocky asteroids around white dwarfs by their periodic thermal emission. *Astrophys. J.* **793,** L43 (2014).
30. Croll, B. *et al.* Multiwavelength observations of the candidate disintegrating sub-Mercury KIC 12557548b. *Astrophys. J.* **786,** 100 (2014).

Supplementary Information is available in the online version of the paper.

Author Contributions A.V. processed and searched the K2 data, identified this system, analysed the K2 data for WD 1145+017 (with help from S.R., D.K., and J.T.W.), processed the MINERVA data, measured radial velocities (with help from W.R.B. and D.W.L.), and was the primary author of the manuscript. S.R. performed the dynamical calculations and dust simulations. W.R.B. obtained and reduced the MMT spectra. P.D. analysed the MMT spectra and SDSS photometry to measure spectroscopic properties. J.A.L. analysed archival photometric measurements and modelled the excess infrared emission. A.B. and D.W.L. obtained and processed the FLWO data. J.I. and D.C. obtained and processed the MEarth data. D.R.C. and C.B. obtained and processed the Keck data. R.A. calculated the systematics insensitive periodogram. L.S. calculated vapour pressures for some minerals with MAGMA. J.A.J., J.E., N.M., R.A.W., and J.T.W. made it possible to use MINERVA. J.A.J. provided scientific leadership.

Author Information The raw K2 data are available at http://archive.stsci.edu/k2/data_search/search.php under the identification number 201563164. The processed K2 data are available at https://archive.stsci.edu/missions/hlsp/k2sff/html/c01/ep201563164.html. We have opted not to make the code used in this work available. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.V. (avanderburg@cfa.harvard.edu).

LETTER

The rise of fully turbulent flow

Dwight Barkley[1], Baofang Song[2,3,4], Vasudevan Mukund[2], Grégoire Lemoult[2], Marc Avila[5] & Björn Hof[2]

Over a century of research into the origin of turbulence in wall-bounded shear flows has resulted in a puzzling picture in which turbulence appears in a variety of different states competing with laminar background flow[1-6]. At moderate flow speeds, turbulence is confined to localized patches; it is only at higher speeds that the entire flow becomes turbulent. The origin of the different states encountered during this transition, the front dynamics of the turbulent regions and the transformation to full turbulence have yet to be explained. By combining experiments, theory and computer simulations, here we uncover a bifurcation scenario that explains the transformation to fully turbulent pipe flow and describe the front dynamics of the different states encountered in the process. Key to resolving this problem is the interpretation of the flow as a bistable system with nonlinear propagation (advection) of turbulent fronts. These findings bridge the gap between our understanding of the onset of turbulence[7] and fully turbulent flows[8,9].

The sudden appearance of localized turbulent patches in an otherwise quiescent flow was first observed by Osborne Reynolds for pipe flow[1] and has since been found to be the starting point of turbulence in most shear flows[2,4,10-15]. In this regime of localized turbulence it is impossible to maintain turbulence over extended regions as it automatically[16,17] reduces to discrete patches, each of approximately the same size. Such patches are called puffs in the context of pipe flow (see Fig. 1a). Puffs can decay, or else split and thereby multiply. For Reynolds numbers (dimensionless flow rates) $R > 2{,}040$, the splitting process outweighs decay, resulting in sustained disordered motion[7]. Although sustained, this turbulence appears only as discrete puffs surrounded by laminar flow (Fig. 1a), and larger clusters of turbulence cannot form[17,18].

At flow rates larger than those sustaining the regime of localized turbulent patches, the situation is fundamentally different: once triggered, turbulence aggressively expands and eliminates all laminar motion (Fig. 1b). The flow is then fully turbulent and only in this state do wall-bounded shear flows have characteristic mean properties such as the Blasius or Prandtl–von Karman friction laws[9]. This rise of fully turbulent flow has remained unexplained, despite the fact that this transformation occurs in virtually all shear flows and generally dominates the dynamics at sufficiently large Reynolds numbers.

A classic diagnostic for the formation of turbulence[2-5,19,20] is the propagation speed of the upstream and downstream fronts of a turbulent patch. We carried out such measurements for pipe and square-duct flow (Fig. 1c), focusing on the regime where turbulence first begins to expand. In both experiments, fluid enters the conduit through a smoothly contracting inlet, which ensures that, without external perturbations, flows are laminar over the Reynolds number range shown in Fig. 1. Turbulence is triggered $120d$ from the inlet (where d is the pipe diameter; see Methods) by a short-duration, localized perturbation. A pressure sensor at the outlet determines the subsequent arrival of first the downstream and then the upstream turbulent–laminar front. Speeds are averaged over many realizations for each R, corresponding to a total travel distance of typically $5 \times 10^4 d$. As an independent verification, speeds in pipe flow were

determined from direct numerical simulations in pipes of length $180d$, with averaging over typically 20 runs.

In both pipe and square-duct flows, initially the speeds of the downstream fronts are indistinguishable from the upstream ones, signalling localized turbulence. For $R \gtrsim 2{,}250$ in pipe flow and $R \gtrsim 2{,}030$ in duct flow, the downstream speed increases with R; these values mark the point where turbulence begins to aggressively invade the surrounding fluid. With further increases in R, the downstream front speeds exhibit complex changes of curvature as a function of R. The spreading of turbulence shows neither a square-root scaling nor an exponent associated with a percolation-type process, as proposed in earlier studies[4,21]; the speed of the downstream spreading exhibits far more complex behaviour than these theories imply.

In a previous theoretical approach[22-24], puffs in pipe flow were categorized as localized excitations, analogous to action potentials in axons, from which the numerous features of puff turbulence were captured. However, in that model, the transition leading to an expanding state is first-order (discontinuous), which does not reflect the observed continuous behaviour at the onset of fully turbulent flow (Fig. 1c). Moreover, this model did not include nonlinear advection, a feature intrinsic to fluid dynamics. We have devised an extended model incorporating an advective nonlinearity that enables us to fully capture the sequence encountered in the transformation to fully turbulent flow. The model is

$$q_t + (u - \zeta)q_x = f(q, u) + Dq_{xx}, \quad u_t + uu_x = \epsilon g(q, u) \qquad (1)$$

where

$$\begin{aligned} f(q, u) &= q(r + u - 2 - (r + 0.1)(q - 1)^2), \\ g(q, u) &= 2 - u + 2q(1 - u) \end{aligned} \qquad (2)$$

and the subscripts denote partial derivatives. The variables q and u depend only on the streamwise coordinate x and time t. q denotes the turbulence level within the flow, which is physically representative of a cross-sectional integration of the turbulent fluctuations. u represents the centreline velocity of the fluid and plays two important roles: it accounts for nonlinear advection in the streamwise direction and captures the physical state of the shear profile, with $u = 2$ corresponding to parabolic flow and $u < 2$ to plug flow. The functions $f(q, u)$ and $g(q, u)$ describe, with minimal nonlinearities, the known interplay between turbulence (the excited state) and the shear profile[3,18,22]. An explicit derivation of these functions from the Navier–Stokes equations has yet to be achieved. The parameter r models the Reynolds number, ζ accounts for the fact that turbulence is advected more slowly than the centreline velocity, D controls the coupling strength of the turbulent patches to the laminar flow (via diffusion) and ϵ sets the timescale ratio between the fast excitation of q and the slow recovery of u following relaminarization; see Methods for details.

To elucidate the core of the transition from localized to expanding excitations, and to identify the different states occurring in the process, we carry out a standard asymptotic analysis[25,26] in the limit of sharp laminar–turbulent fronts ($\epsilon \to 0$). Three distinct turbulent structures are predicted: a localized state (Fig. 2a), an asymmetric expanding state
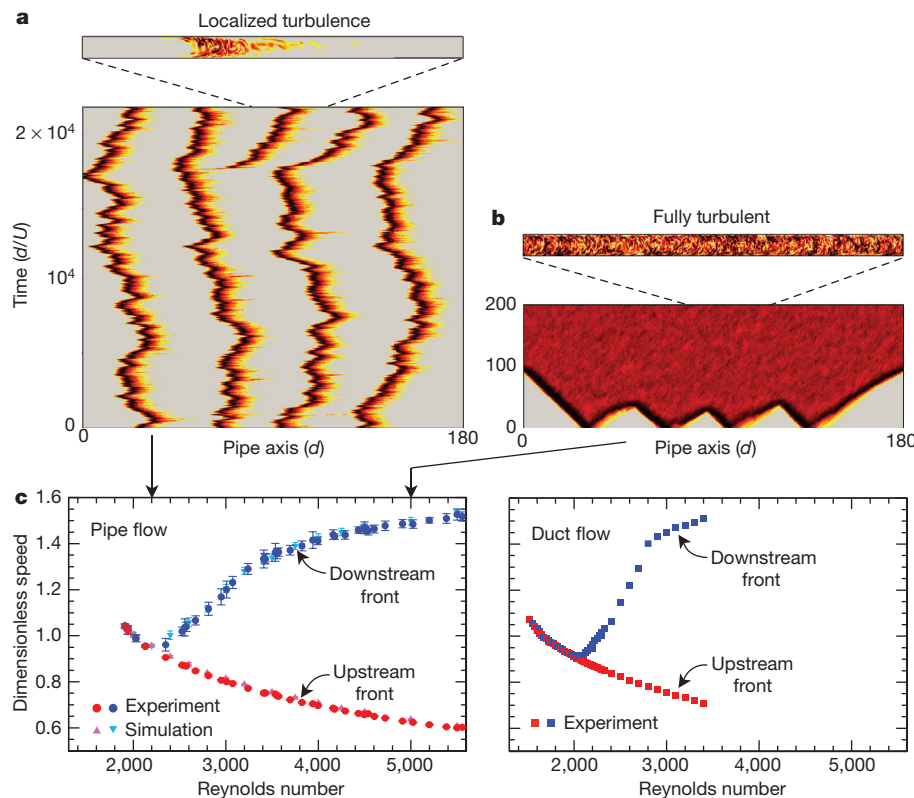
Figure 1 | Localized and fully turbulent flow.
a, **b**, Numerical simulations of pipe flow illustrate the distinction between localized turbulence at $R = 2{,}200$ (**a**) and fully turbulent flow at $R = 5{,}000$ (**b**). In each case, the flow is initially seeded with localized turbulent patches and the subsequent evolution is visualized via space–time plots in a reference frame co-moving with the structures. Colours indicate the value of $\sqrt{u_r^2 + u_\theta^2}$, where $\boldsymbol{u} = (u_r, u_\theta, u_x)$ is the velocity vector at the given point, expressed in cylindrical coordinates. Cross-sections of instantaneous flow within the pipe are shown above the plots to further highlight the distinction between the two regimes; a $35d$ section is shown with the vertical direction (the pipe cross-section) stretched by a factor of two. Although the protocol used here is seeding the flow with localized patches of turbulence, the fundamental distinction between localized and fully turbulent flow is independent of how turbulence is triggered[16]. **c**, Speeds of turbulent–laminar fronts as a function of Reynolds number for pipe flow and duct flow. Error bars give 95% confidence interval for average values (symbols). Speeds are nondimensionalized by the mean streamwise velocity U. A speed difference between the upstream and downstream fronts corresponds to expanding turbulence. The arrows at the top of the left panel indicate the Reynolds numbers at which the simulations in **a** and **b** were performed.

(Fig. 2b) and a symmetric expanding state (Fig. 2c). The essence of each state is seen in the local phase plane (Fig. 2d–f). Equilibrium points are located at the intersections of the q and u nullclines (curves where time derivatives of u and q are zero). For low values of r (Fig. 2d) the only equilibrium is $(u = 2, q = 0)$, corresponding to parabolic laminar flow. Nevertheless, the system can be excited locally; when perturbed, the state jumps to the upper branch q^+. This forms the upstream laminar-to-turbulent front. On the upper branch, $\dot{u} < 0$ (where the overdot indicates differentiation with respect to time) and u decreases to a point where turbulence is not maintained and the system jumps back to $q = 0$, forming the downstream front. The downstream front follows the upstream one at a fixed distance, thus creating a localized excitation: a puff in pipe flow analogous to an action potential in excitable media[24–26].

For larger values of r, a second stable equilibrium appears (uppermost intersection of the nullclines in Fig. 2e, f) and the system is now bistable. Here, fully turbulent flow begins to arise. The downstream front lags the upstream front, giving rise to a growing turbulent region between the fronts. Initially the expansion is asymmetric and the spreading rate is modest (Fig. 2b, e); the fronts themselves are not very different in appearance from those of the localized state. The downstream front occurs at $u < 2$ and is formed by a drop directly from the upper equilibrium $q = q^+$ to $q = 0$ (Fig. 2e). We refer to this as the 'weak front state'. For larger r the weak front becomes unstable, giving rise to the final state, a much more rapidly expanding 'strong front state' (Fig. 2c, f). The strong downstream front occurs at $u = 2$ and is the mirror image of the upstream front. As seen in Fig. 2c, the value of q increases above q^+ just before the drop to $q = 0$ at the downstream front. The downstream speed is opposite to the upstream speed with respect to what we term the 'neutral speed'.

Before comparing the model to the experimental data, we discuss features of the front-speed scaling that are intrinsic to this model. Figure 2g shows front speeds of the three states. (From the asymptotic analysis presented in the Methods, the front speeds explicitly scale as $\sqrt{D}$; the results in Fig. 2 are for $D = 0.13$.) Starting at low r, excitations

are strictly localized and their speed monotonically decreases with r (red curve in Fig. 2g). Expanding turbulence is first encountered when this curve intersects the weak-front curve (green in Fig. 2g). The turbulent state (upper fixed point in Fig. 2e, f) bifurcates at lower r, but initially the downstream speed is smaller than the upstream one, resulting in a contraction back to a localized excitation. Thus onset of bistability and the expansion do not coincide, masking the transition and resulting in a non-standard front speed scaling (in contrast to the case without nonlinear advection shown in Extended Data Fig. 1a). The strong front (blue in Fig. 2g) is stable at slightly higher r (solid portion of the curve) and is perfectly symmetric to the downstream front (red in Fig. 2g) about the neutral speed. In the asymptotic limit ($\epsilon \to 0$), weak and strong fronts co-exist over a range of r, but for finite ϵ the front speed continuously varies from a weak to increasingly strong front (solid black curve in Fig. 2g). During this adjustment the front speed exhibits two curvature changes. This, together with the eventual approach to the upper branch of the parabola, is a distinct signature of the scenario described by this model.

Using the theoretical model as a guide, we combine the measured front speeds from pipe and duct flow and compare them directly with theory (Fig. 3a). Initially, at lower values of R, turbulent excitations are localized (as illustrated for duct flow in Fig. 3b and pipe flow in Fig. 3e) and the front speed data from both flows agree very well with the parabolic scaling predicted by the model asymptotics (solid red curve in Fig. 3a). At $R \approx 2{,}250$ in pipe flow and $R \approx 2{,}030$ in duct flow, expansion begins with the formation of the weak downstream front (illustrated for duct and pipe flow in Fig. 3c, f, respectively). Although upstream fronts of both data sets continue to follow the simple asymptotic form, the weak downstream fronts do not display the same scaling. Nevertheless, with appropriate choices of the parameters ζ and ϵ, the model precisely captures the two curvature changes (solid black curves in Fig. 3a) encountered as each flow continuously adjusts from the weak front (green dashed line in Fig. 3a) to the strong front (blue dashed line in Fig. 3a), corresponding to the emergence of the final strong front state (Fig. 3d, g). As the downstream front
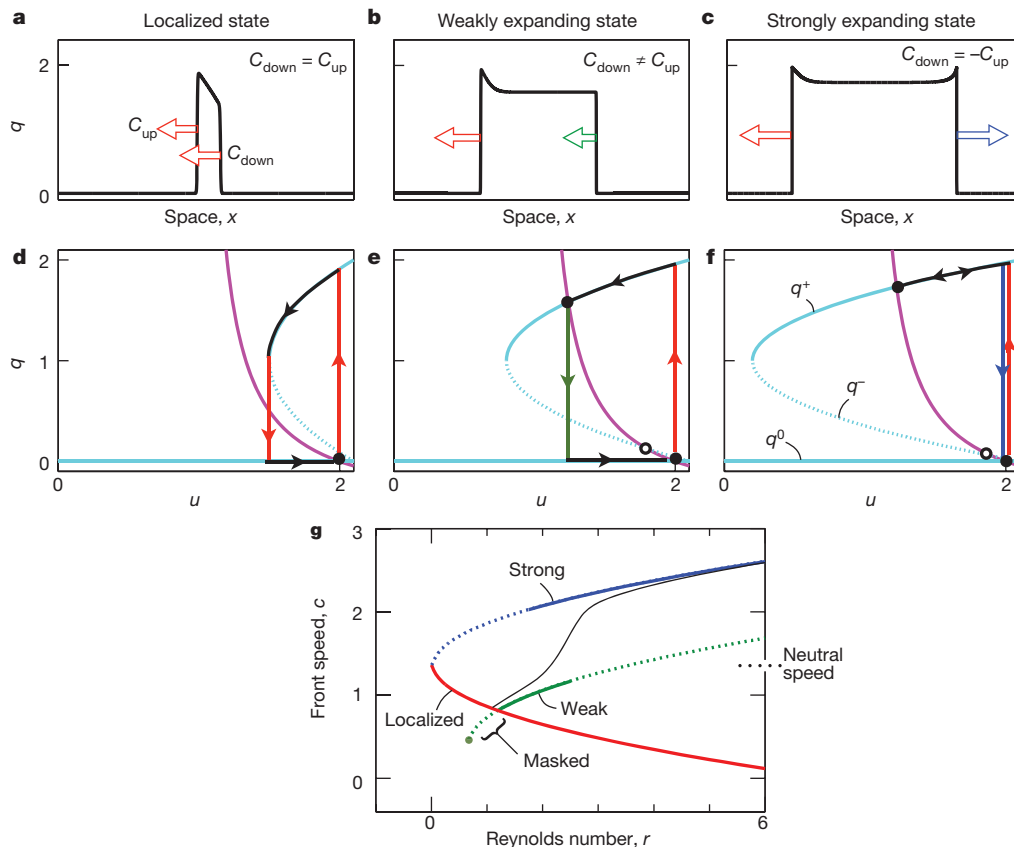
Figure 2 | Model predictions in the asymptotic limit of sharp laminar–turbulent fronts. a–c, Three distinct types of predicted states. C_{up} and C_{down} are the upstream and downstream front speeds. **d–f**, Corresponding states viewed in the local phase plane, with arrows indicating increasing x (not time). The q nullcline ($f(q,u) = 0$; cyan, for clarity labelled only in **f**) has three branches because $f(q,u)$ is cubic in q: two stable branches, laminar $q^0 = 0$ and the upper $q = q^+$ branch, and an unstable branch $q = q^-$ separating the two stable branches. The u nullcline ($g(q,u) = 0$; magenta) describes the decrease in the centreline velocity in the presence of turbulence and its recovery in the absence of turbulence. Fronts are formed when the system jumps between stable branches of the q nullcline. In all cases, the upstream front is a transition from laminar flow (the equilibrium at $u = 2$, $q = 0$, indicated by a filled black circle) to the upper branch $q = q^+$. (These fronts are shown as red lines with up arrows in **d–f**; the corresponding speeds are indicated by red arrows in **a–c**.) The cases are distinguished by the downstream front. In **a** and **d** ($r = 0.5$), the system is excitable and the downstream transition, from q^+ to q^0 (red line with down arrow in **d**), is unrestricted by the upper branch and the speed (also red in **a**) will be selected to match the upstream front $C_{down} = C_{up}$, yielding

localized turbulence. In **b** and **e** ($r = 1.2$), the system has become bistable with the formation of an upper-branch steady state (indicated by the second filled black circle). There is also an unstable fixed point, indicated by the open circle. Evolution on the upper branch is restricted by this state so the downstream front speed may no longer be able to match the upstream front speed: in general $C_{down} \neq C_{up}$. Consequently, the turbulent patch expands. In **c** and **f** ($r = 1.8$), the upstream and downstream fronts have the same character, but move in opposite directions, $C_{down} = -C_{up}$, in a reference frame moving at the neutral speed. We refer to the downstream fronts in **b** and **e** as 'weak fronts' (shown in green) and those in **c** and **f** as 'strong fronts' (shown in blue). **g**, Front speeds as a function of model Reynolds number r. Upstream and localized downstream speeds (red), weak front speeds (green), and strong front speeds (blue) are from equations (10) and (11) in Methods, with solid lines indicating stable fronts as $\epsilon \to 0$. The nominal critical point for the onset of fully turbulent flow is masked. The neutral speed is the speed about which the upstream and strong downstream front speeds are symmetric. At finite ϵ, the transition from weak to strong scaling is continuous (black curve).

approaches the scaling given by the strong-front asymptotics, its speed forms a parabola with the upstream front speed, a feature overlooked in previous studies.

Weak fronts move more slowly than the bulk advection velocity of turbulence; once the downstream front speed exceeds the bulk advection velocity, the front switches to a strong front. At that point, a turbulent patch invades (nearly) fully recovered laminar flow at the downstream front, in much the same way that turbulence invades fully recovered laminar flow at the upstream front. This produces the symmetry between the upstream and strong downstream fronts.

There are two features of pipe and duct turbulence that the model does not capture. Both originate from stochastic fluctuations within turbulence and are most prevalent when turbulence first begins to expand (Fig. 3c, f). Fronts fluctuate, especially the downstream front, and it is common for the system to sometimes exhibit a strong and sometimes a weak downstream front. The bifurcation scenario predicted by the model is only recovered in average quantities. Likewise,

turbulence for $2{,}250 \lesssim R \lesssim 3{,}000$ is not always uniform, but commonly contains intermittent laminar pockets[16,27].

The simplicity of the model permits investigation of new phenomena associated with fully turbulent flow. In the model the creation of extended turbulent regions hinges on the upper intersection of the q and u nullclines, and by manipulating u this fixed point can be destroyed (see Methods). Likewise for pipe flow an analogous profile manipulation leads to a reverse transition. As demonstrated in the Methods, fully turbulent flow is eliminated and only localized excitations remain, offering a very simple and robust way to control turbulence and to reduce frictional drag.

Although much progress has been made in our understanding of how turbulence in wall-bounded flows is formed from unstable invariant solutions[28–30] at moderate R, little to no progress has been made in connecting this transitional regime to studies of high-R turbulence. Explaining the origin of the fully turbulent state is a decisive step towards connecting these regimes and paves the way for a bottom-up approach to turbulence.
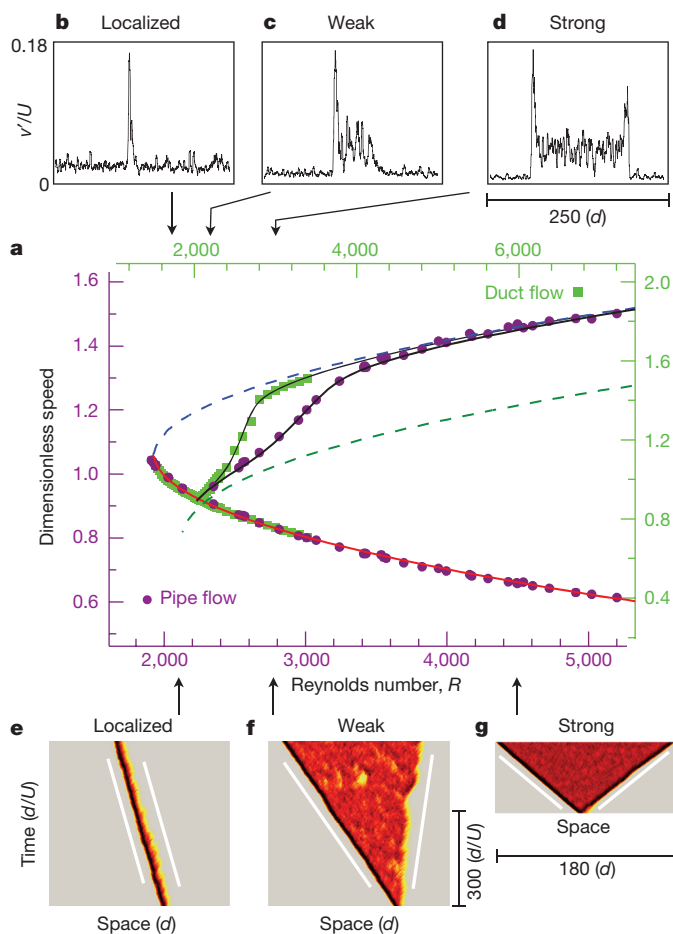
Figure 3 | The rise of fully turbulent flow. a, Front speeds as a function of Reynolds number for pipe and duct flow. Points are experimental results from Fig. 1c. Red, blue and dark green curves are front speeds in the asymptotic limit of sharp fronts in q (as in Fig. 2g). The only model parameter used to fit these curves is $D = 0.13$. Black curves are the downstream front speed at finite front width ($\epsilon = 0.2$, $\zeta = 0.79$ for pipe flow and $\epsilon = 0.11$, $\zeta = 0.56$ for duct flow). The distinct weak and strong asymptotic branches (dashed) form the skeleton for the formation of fully turbulent flow, while at finite front width the model captures the complex behaviour of front speeds as a smooth switching between the asymptotic branches. **b–d**, Cross-stream velocity fluctuations v'/U for the three front states in a square duct: localized puff ($R = 1,700$), expanding turbulence with a weak downstream front ($R = 2,300$) and the strong front state ($R = 3,000$), which exhibits the characteristic energy overshoot at the downstream edge[5,20] (the arrows to **a** indicate the Reynolds number to which **b–d** correspond). **e–g**, Space–time plots from simulations of pipe flow at $R = 2,000$, $R = 2,800$ and $R = 4,500$, respectively (as indicated by the arrows to **a**). $\sqrt{u_r^2 + u_\theta^2}$ is plotted in the reference frame moving at the neutral speed. White lines indicate front speeds from the model converted to physical units. At $R = 2,000$, turbulence is localized with equal upstream and downstream front speeds. At $R = 4,500$, turbulence expands with a strong downstream front and the long-time flow is fully turbulent. The upstream and downstream fronts have the same character (compare with the symmetric overshoot in Fig. 2d) and the spreading is symmetric in the neutral reference frame. At $R = 2,800$, the downstream front moves at a speed between the weak and strong branches and exhibits some characteristics of both fronts as it fluctuates. This, as well as the intermittent laminar patches appearing within the turbulent flow, is typical of turbulence as fully turbulent flow first arises.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Reynolds, O. An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Phil. Trans. R. Soc. Lond. A* **174**, 935–982 (1883).

2. Coles, D. Interfaces and intermittency in turbulent shear flow. In *Mécanique de la Turbulence* 229 (Éditions du Centre National de la Recherche Scientifique, 1962).
3. Wygnanski, I. J. & Champagne, F. H. On transition in a pipe. Part 1. The origin of puffs and slugs and the flow in a turbulent slug. *J. Fluid Mech.* **59**, 281–335 (1973).
4. Sreenivasan, K. R. & Ramshankar, R. Transition intermittency in open flows, and intermittency routes to chaos. *Physica D* **23**, 246–258 (1986).
5. Nishi, M., Ünsal, B., Durst, F. & Biswas, G. Laminar-to-turbulent transition of pipe flows through puffs and slugs. *J. Fluid Mech.* **614**, 425–446 (2008).
6. Mullin, T. Experimental studies of transition to turbulence in a pipe. *Annu. Rev. Fluid Mech.* **43**, 1–24 (2011).
7. Avila, K. *et al.* The onset of turbulence in pipe flow. *Science* **333**, 192–196 (2011).
8. Pope, S. B. *Turbulent Flows* Ch. 7 264–332 (Cambridge Univ. Press, 2000).
9. Schlichting, H. & Gersten, K. *Boundary-Layer Theory* 8th edn (Springer, 2000).
10. Emmons, H. W. The laminar-turbulent transition in a boundary layer—part I. *J. Aeronaut. Sci.* **18**, 490–498 (1951).
11. Coles, D. Transition in circular Couette flow. *J. Fluid Mech.* **21**, 385–425 (1965).
12. Lundbladh, A. & Johansson, A. V. Direct simulation of turbulent spots in plane Couette flow. *J. Fluid Mech.* **229**, 499–516 (1991).
13. Tillmark, N. & Alfredsson, P. H. Experiments on transition in plane Couette flow. *J. Fluid Mech.* **235**, 89–102 (1992).
14. Cros, A. & Le Gal, P. Spatiotemporal intermittency in the torsional Couette flow between a rotating and a stationary disk. *Phys. Fluids* **14**, 3755–3765 (2002).
15. Lemoult, G., Aider, J.-L. & Wesfreid, J. E. Turbulent spots in a channel: large-scale flow and self-sustainability. *J. Fluid Mech.* **731**, R1 (2013).
16. Moxey, D. & Barkley, D. Distinct large-scale turbulent-laminar states in transitional pipe flow. *Proc. Natl Acad. Sci. USA* **107**, 8091–8096 (2010).
17. Samanta, D., de Lozar, A. & Hof, B. Experimental investigation of laminar turbulent intermittency in pipe flow. *J. Fluid Mech.* **681**, 193–204 (2011).
18. Hof, B., De Lozar, A., Avila, M., Tu, X. & Schneider, T. Eliminating turbulence in spatially intermittent flows. *Science* **327**, 1491–1494 (2010).
19. Lindgren, E. R. The transition process and other phenomena in viscous flow. *Ark. Fys.* **12** (1957).
20. Duguet, Y., Willis, A. P. & Kerswell, R. R. Slug genesis in cylindrical pipe flow. *J. Fluid Mech.* **663**, 180–208 (2010).
21. Sipos, M. & Goldenfeld, N. Directed percolation describes lifetime and growth of turbulent puffs and slugs. *Phys. Rev. E* **84**, 035304 (2011).
22. Barkley, D. Simplifying the complexity of pipe flow. *Phys. Rev. E* **84**, 016309 (2011).
23. Barkley, D. Modeling the transition to turbulence in shear flows. *J. Phys. Conf. Ser.* **318**, 032001 (2011).
24. Barkley, D. Pipe flow as an excitable medium. *Rev. Cub. Fis.* **29**, 1E27 (2012).
25. Rinzel, J. & Terman, D. Propagation phenomena in a bistable reaction-diffusion system. *SIAM J. Appl. Math.* **42**, 1111–1137 (1982).
26. Tyson, J. & Keener, J. Singular perturbation theory of traveling waves in excitable media (a review). *Physica D* **32**, 327–361 (1988).
27. Avila, M. & Hof, B. Nature of laminar-turbulence intermittency in shear flows. *Phys. Rev. E* **87**, 063012 (2013).
28. Eckhardt, B., Schneider, T. M., Hof, B. & Westerweel, J. Turbulence transition in pipe flow. *Annu. Rev. Fluid Mech.* **39**, 447–468 (2007).
29. Gibson, J., Halcrow, J. & Cvitanović, P. Visualizing the geometry of state space in plane Couette flow. *J. Fluid Mech.* **611**, 107–130 (2008).
30. Kawahara, G., Uhlmann, M. & van Veen, L. The significance of simple invariant solutions in turbulent flows. *Annu. Rev. Fluid Mech.* **44**, 203–225 (2012).

Author Contributions V.M., G.L. and B.H. designed and performed the experiments and analysed the experimental results. B.S. and M.A. designed and performed the computer simulations of the Navier–Stokes equations. B.S., M.A. and B.H. analysed the numerical results. B.S. generated the corresponding visualizations. D.B. performed the theoretical analysis. D.B., B.S., V.M., G.L., M.A. and B.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (D.Barkley@warwick.ac.uk, for theoretical aspects) and B.H. (bhof@ist.ac.at, for experimental aspects).

METHODS

Speed measurements. Speeds of laminar–turbulent fronts were measured in experiments and highly resolved computer simulations. In both cases, long observation times were necessary to average out stochastic fluctuations that, although intrinsic to turbulence, may disguise the underlying transition scenario. All measured speeds are nondimensionalized by the mean streamwise velocity U. Times are reported in units of d/U for pipe flow and h/U for duct flow, where d is the pipe diameter and h is the duct width. The corresponding Reynolds numbers for the two flows are $R = Ud/v$ and $R = Uh/v$, where v is the kinematic viscosity.

Pipe experiments. Experiments were carried out in a pipe with a diameter $d = 10$ mm (± 0.01 mm) and a length of $1,500d$. The 15-m-long pipe was assembled on a straight aluminium base and made of precision bore glass tubes with lengths of 1–1.2 m. Customized connectors made from perspex allowed an accurate fit of the pipe segments. A specially made pipe inlet consisting of several meshes and a smooth convergence from a 100-mm-wide section to the 10-mm pipe was used to avoid inlet disturbances and eddie formation (see ref. 17 for details). In this way, the water flow could be held laminar for $R > 8,000$.

The laminar flow was left to develop its parabolic velocity profile over a length of $200d$. At this downstream location, the flow was perturbed by an impulsive jet of water injected (for 10 ms) through a 1-mm hole in the pipe wall. The perturbed flow was left to develop into a turbulent patch over the next $250d$ and at this location ($450d$ from the inlet), a pressure sensor recorded the arrival of the upstream and downstream laminar–turbulent interfaces. A second sensor was located a further $1,000d$ downstream ($50d$ upstream of the pipe exit), once again determining the arrival of the interfaces so that the average interface speed over the intermediate stretch of $1,000d$ was measured. At each Reynolds number, the measurement of the interface velocity was repeated 10 times.

The flow was gravity driven from a reservoir at a fixed height above the pipe exit. Because the turbulent fraction in the pipe is increasing over the course of a measurement, the overall drag in the pipe also increases (turbulent flow has a higher skin friction than does laminar flow). This unavoidably leads to a drop in the flow rate (and hence R) during a measurement. To minimize this effect, a large reservoir height was chosen; in this case 23 m above the pipe exit. A precision valve positioned directly in front of the pipe inlet was used to adjust the flowrate and hence to select R. For the Reynolds-number regime investigated here ($R < 6,000$), the total pressure drop across the pipe is much smaller than the 23-m water head, and most of the pressure drop occurs across the valve. The increase in drag caused by the expansion of turbulence is only a small fraction ($<0.5\%$ of the overall pressure drop) and hence, even at the highest Reynolds numbers, investigated flow rates were constant to within $<0.5\%$ throughout the measurement.

Duct experiments. Experiments were carried out in a square duct with width $h = 5$ mm and a length of $1,200h$ (6 m). The duct was made of eight perspex sections precisely machined to an accuracy of ± 0.01 mm. They were assembled and mounted straight together on an aluminium frame. A well-designed entrance section consisting of a honeycomb and a convergent section, with an area ratio of 25, allowed the flow to remain laminar up to at least $R = 5,000$.

The flow was gravity driven from a reservoir at a fixed height and water was used as the working fluid. Analogous to the pipe experiment, a precision valve was positioned directly in front of the duct and was used to set the flowrate. The pressure drop across the valve was considerably larger than that across the pipe. The temperature of the water was controlled by means of a heat exchanger that the water had to pass before entering the pipe. Overall, an accuracy in R of better than 0.5% was achieved for the investigated Reynolds number regime ($R < 6,000$).

The flow was perturbed by injecting water through a 0.5-mm hole drilled in one wall of the duct, $120h$ downstream from the inlet. The duration of the perturbation (t) was varied with R so that in dimensionless units it corresponded to 5 ($t/h/U$). The evolution of the perturbation was then monitored at five locations where the pressure was recorded. The pressure sensors were positioned at $100h$, $400h$, $600h$, $800h$ and $1,000h$ downstream of the perturbation point. Sensors measured the pressure difference over $10h$ along the duct. The arrival times of both interfaces were detected at each location and the overall speeds were determined by a linear fit. For each R, we averaged the measurement over at least 50 realizations.

Numerical simulations. We consider the motion of incompressible fluid driven through a circular pipe with a fixed mass flux. Normalizing lengths with the diameter d and velocities with the mean velocity U, the Navier–Stokes equations read

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + \frac{1}{R}\Delta\boldsymbol{u}, \quad \nabla \cdot \boldsymbol{u} = 0$$

where $\boldsymbol{u}$ is the velocity of the fluid and p is the pressure. These equations were solved in cylindrical coordinates (r, θ, x) using a code developed by A. P. Willis[31], which uses a spectral finite-difference method with no-slip boundary conditions at the pipe wall, $\boldsymbol{u}(1/2, \theta, x, t) = \boldsymbol{0}$ and periodicity in the axial direction. The pressure term was eliminated from the equations by using a toroidal–poloidal potential formulation of the velocity field, in which the velocity is represented by toroidal ψ and poloidal potentials ϕ, such that $\boldsymbol{u} = \nabla \times (\psi\hat{\boldsymbol{x}}) + \nabla \times \nabla \times (\phi\hat{\boldsymbol{x}})$.

After projecting the curl and double curl of the Navier–Stokes equations onto the x axis, a set of equations for the potentials ψ and ϕ is obtained. A difficulty, due to the coupled boundary conditions on the potentials, is solved with an influence-matrix method. In the radial direction, spatial discretization is performed using a finite-difference method with a 9-point stencil. Assuming periodicity in azimuthal and axial directions, the potentials are expanded in Fourier modes

$$A(r, \theta, x, t) = \sum_{k=-K}^{K} \sum_{m=-M}^{M} \hat{A}_{k,m}(r, t)e^{i\alpha kx + im\theta} \tag{3}$$

where αk and m are the wavenumbers of the modes in the axial and azimuthal directions respectively, $2\pi/\alpha$ fixes the pipe length L_x, and $\hat{A}_{k,m}$ is the complex Fourier coefficient of mode (k, m). The time-dependent equations are integrated in time using a second-order predictor–corrector scheme with a dynamic timestep size, which is controlled using information from a Crank–Nicolson corrector step. The nonlinear term is evaluated using a pseudo-spectral technique with the de-aliasing $\frac{3}{2}$-rule. Using the expansion in equation (3), the resultant linear differential equations for the potentials ψ and ϕ decouple for each (k, m) mode. This linear system is solved using LU decompositions of the resultant banded matrices; see ref. 31 for more details of the formulation and solution.

Initial conditions were prepared at $R = 2,000$ in $133d$ and $180d$ pipes for simulations at $R > 2,000$. At low $R \approx 2,000$, puff-splitting is extremely unlikely[7] and puffs remain approximately constant in length (about $20d$) as they travel downstream along the pipe. Hence, simulations at $R = 1,910$, $1,920$, $2,000$ were carried out in a shorter $24\pi \approx 75d$ pipe, with initial conditions prepared at $R = 1,950$. The lengths of the pipes and numerical resolutions used at each Reynolds number are listed in Extended Data Table 1.

The fronts were detected by setting an appropriate cut-off. Here, the local intensity was computed as

$$\int\int (u_r^2 + u_\theta^2)r\,\mathrm{d}r\,\mathrm{d}\theta$$

and a cut-off of 5×10^{-4} was chosen for all the simulations to determine the position of the laminar–turbulent fronts. We tested different cut-off values and found that the front speed was insensitive to the chosen value.

The expansion speed of the downstream front was found to accelerate substantially during the initial stages of the simulation. To obtain the asymptotic value of the speed, we determined the length of the turbulent region L_0 beyond which the speed statistics become length-independent. We found that for $R < 4,000$, $L_0 > 60d$ was sufficient, whereas for $R \geq 4,000$, $L_0 > 100d$ was required. This is the reason why very long pipes were used, as reported in Extended Data Table 1. At each R, the speed was determined by computing $(x_{\mathrm{end}} - x_0)/(t_{\mathrm{end}} - t_0)$ for each run and then averaging over a total of 20 runs. The initial time t_0 corresponds to the time at which the turbulent region has reached the length L_0.

Model details. The model is a two-component system of advection-reaction-diffusion equations

$$\frac{\partial q}{\partial t} + (u - \zeta)\frac{\partial q}{\partial x} = f(q, u) + D\frac{\partial^2 q}{\partial x^2}, \quad \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \epsilon g(q, u) \tag{4}$$

where q represents the level of turbulent fluctuations and u the axial velocity at the centreline. The nonlinear reaction functions $f(q, u)$ and $g(q, u)$ are

$$f(q, u) = q(r + u - 2 - (r + 0.1)(q - 1)^2), \quad g(q, u) = 2 - u + 2q(1 - u)$$

where the parameter r corresponds to the model Reynolds number.

The model and the role of the fitting parameters (D, ζ and ϵ) are most easily understood by first considering the equations in the absence of spatial derivatives. In this case, the model reduces to the ordinary differential equations (ODEs)

$$\dot{q} = f(q, u), \quad \dot{u} = \epsilon g(q, u)$$

where the overdot indicates differentiation with respect to time. These ODEs are the core of the model as they describe the interaction between the turbulent fluctuations q and axial velocity u locally in space. The functional forms are designed to qualitatively capture the well-established physics of this interaction[2,3] with minimal nonlinearities. (In a previous approach[22], the variable u corresponded to the axial velocity of pipe flow in the frame of reference moving at the mean or bulk velocity U; here, u corresponds to velocity in the lab frame so that $u = 2$ for laminar flow.)

The nullclines for the ODEs are $f(q, u) = 0$ and $g(q, u) = 0$. For all parameter values, these nullclines intersect at the fixed point ($u = 2$, $q = 0$) corresponding to laminar, Hagen–Poiseuille flow. ϵ sets the ratio of the timescale of u relative to q.

(Previously[22], two parameters ϵ_1 and ϵ_2 appeared in the model; here we have simplified the model to a single timescale ratio ϵ, where $\epsilon_1 \equiv \epsilon$ and $\epsilon_2 \equiv 2\epsilon$.)

Now consider the full model equations. In addition to the local terms given by $f(q,u)$ and $g(q,u)$, the model has first and second spatial derivatives. The first-derivative terms account for nonlinear advection in the streamwise direction. For the u equation, we use the advective nonlinearity that follows directly from the Navier–Stokes equations. The parameter ζ accounts for diminished advection of q in comparison with the centreline velocity u. The streamwise velocity is maximal on the centreline and the turbulent field is not advected at this speed. We simulated turbulent flow in short pipes ($L = 12d$) and verified that turbulent structures are advected considerably more slowly than is the centreline velocity. This effect leads to complex processes in the pipe cross-section. We include in the model the simplest term that can describe the diminished advection. (Previously[22], the model contained only linear advection; the fixed difference in the advection of the q and u fields was expressed by an additional first-derivative term on the right-hand side of the u equation, which effectively corresponded to $\zeta = 1$ in the current model.) We describe the importance of the parameter ζ after we derive expressions for front speeds in the model.

The diffusive term in equation (4) accounts for the processes by which a region of turbulent flow couples to, and thereby excites, adjacent laminar flow. The physical processes involved are complex and not fully understood[3,7,16,20,32,33]. However, the second-derivative is the most natural choice for modelling such a coupling. The coupling strength or diffusion coefficient D is the final model parameter.

Asymptotic analysis. The asymptotic analysis follows very closely that of ref. 26. Let the three roots of $f(q,u)$ be denoted $q^0, q^{\pm}$. The laminar branch is $q^0 = 0$ for all u and r, whereas the upper and lower branches $q^{\pm}$ are functions of u and r. The laminar q^0 and upper q^+ branches are stable. For small ϵ, the dynamics of the system separate into slow regions and fast front regions. In the slow regions, the system is 'slaved' to one of the stable branches (slow manifolds) and u evolves on a slow scale; for example, along the upper branch q^+

$$q = q^+(u), \qquad \frac{\partial u}{\partial t'} + u \frac{\partial u}{\partial x'} = g(q^+(u), u)$$

where $x' = \epsilon x$ and $t' = \epsilon t$ are the slow scales.

In the fast regions, fronts in q are formed as the system transitions between the stable branches: from q^0 to q^+ as x increases for an upstream front and from q^+ to q^0 for a downstream front. Let c denote the speed of the front and consider a frame of reference moving at speed c. We set the location of the now stationary front at $x = 0$ and work in an inner (stretched) variable $x/\sqrt{D}$. To leading order in ϵ, the equations in the stretched coordinate become

$$q'' + sq' + f(q,u) = 0 \qquad (5)$$
$$u' = 0 \qquad (6)$$

where

$$s \equiv \frac{c - (u_f - \zeta)}{\sqrt{D}}$$

and the prime denotes differentiation with respect to the stretched variable. From equation (6), u is constant to leading order across a front; we denote this constant value u_f. Equation (5) must be solved subject to boundary conditions, which for a downstream front are

$$q(-\infty) = q^+(u_f), \quad q(+\infty) = q^0 \qquad (7)$$

For an upstream front, the boundary conditions are reversed, but this can be accounted for by a change of sign of s in equation (5). These inner solutions determine the shape of the front. In the original length scale x, the front thickness scales as $\sqrt{D}$.

In summary, the speed of a front at a given value of $u = u_f$ is found by solving

$$q'' + sq' + f(q, u_f) = 0 \qquad (8)$$

subject to the boundary conditions in equation (7). This gives a value of s that is a function of both u_f and r, and which we denote by $s(u_f, r)$. From this the front speed is

$$c = u_f - \zeta \pm \sqrt{D} s(u_f, r) \qquad (9)$$

with $+$ for a downstream front and $-$ for an upstream front. For the strong downstream front and all upstream fronts, $u_f = 2$. Hence their speeds are

$$c = 2 - \zeta \pm \sqrt{D} s(2, r) \qquad (10)$$

For the weak downstream front, $u_f = u_{ss}$ where u_{ss} is the upper-branch steady state. Hence

$$c = u_{ss} - \zeta + \sqrt{D} s(u_{ss}, r) \qquad (11)$$

Extended Data Figure 1 shows model front speeds as a function of model Reynolds number (Extended Data Fig. 1b is the same as Fig. 2g); speeds are from equations (10) and (11).

The neutral speed in the model is $c = 2 - \zeta$. This follows immediately from equation (10) where one can see that the upstream speed (minus sign) and strong downstream speed (plus sign) are symmetric with respect to $2 - \zeta$. This is the advection speed of turbulence in the absence of front dynamics due to transitions between laminar and turbulent flow. Without the parameter ζ, the neutral speed would be the maximum centreline velocity. This is neither consistent with the observed neutral speed, nor is it reasonable that turbulent structures would be advected at the maximum speed found in the flow.

Extended Data Figure 1a shows front speeds without the inclusion of advection terms (first derivatives in x) in the model equations. Without these terms the front speeds become

$$c = \pm \sqrt{D} s(2, r)$$

for the strong downstream front and all upstream fronts, and

$$c = \sqrt{D} s(u_{ss}, r)$$

for the weak downstream front. The transition to expanding turbulence is discontinuous. Including linear advection (as was done previously[22]) will result in an overall shift in all front speeds, and can affect the asymptotic stability of branches, but will not change the discontinuous nature of the transition.

This highlights the role of nonlinear advection in the bifurcation scenario: without the physical effect of nonlinear advection, the weak front branch has a distinct critical point and the transition to expanding turbulence is first-order (discontinuous).

In Fig. 2a–c, solutions $q(x)$ are obtained from the full model equations (4) with $\epsilon = 0.002$, which is sufficiently small that these solutions are visually close approximations to the $\epsilon \to 0$ limit. Figure 2d–f, shows the nullclines for the cases shown in Fig. 2a–2c; however, the trajectories in the phase portraits are sketches (with the fronts coloured for clarity): even at this small ϵ, the jumps between the branches of q are not completely vertical in the phase plane.

A further calculation determines the stability of the asymptotic branches (D.B., manuscript in preparation). The result is that the weak downstream front is stable in the asymptotic limit ($\epsilon \to 0$) if $c < u_f = u_{ss}$, whereas the strong downstream front is stable in the asymptotic limit if $c > u_f = 2$. These criteria determine the stable portions of the branches (plotted as solid) in Fig. 2g.

There are many documented exact coherent structures in pipe flow. Most of these are spatially extended, in the form of travelling waves[28–30,34], but spatially localized states have also been found[35,36]. The model captures these states in a minimal way. The fixed points $q^{\pm}$ (one stable and one unstable) arising as the model transitions to bistability can be viewed as upper and lower branches of spatially extended travelling-wave solutions. The cubic nonlinearity in $f(q,u)$ is the minimum requirement for this separation into upper and lower branch states. The model also has localized states (puffs) and, importantly, unstable small-amplitude localized solutions (not discussed here; see refs 22, 25, 26) corresponding to edge states, both in the puff regime and in the fully turbulent regime.

Finally, we comment on what takes place at the critical point where the system first becomes bistable. As with all material in this section, the discussion follows closely refs 25, 26. Extended Data Figure 2 illustrates solutions to the boundary value problem in equation (8) in the case of a downstream front. For a fixed value of r, the eigenvalue s and solution q depend on u_f, as do the boundary conditions in equation (7).

Downstream fronts are heteroclinic connections from q^+ to q^0, where 'time' in the phase plane corresponds to space (in the reference frame co-moving at the front speed). The phase plane is two-dimensional, coordinates q and q', because equation (8) is the second-order ODE. As illustrated in Extended Data Fig. 2c, for generic u_f, both q^+ and q^0 are hyperbolic fixed points (saddles) in the phase plane and a heteroclinic connection exists only for a unique value of s. This determines s as a function of u_f as shown by the bold curve in Extended Data Fig. 2a. However, when u_f is such that $q^+ = q^-$ (at the nose of the q nullcline), the upper fixed point ($q = q^+ = q^-$) is no longer hyperbolic and there exist infinitely many heteroclinic connections from $q^+ = q^-$ to q^0, and hence infinitely many possible values of s. These appear as the thin line in Extended Data Fig. 2a.

As the parameter r is varied (as in Fig. 2), the nullclines vary. The critical point is where the upper-branch steady state occurs at the limit point of the q nullcline, that is the upper fixed point is at $q^+ = q^-$. For r smaller than this value, the downstream front speed can take any of an infinite range of values because the downstream front occurs at $u_f = u_c$. For a puff, this infinite range of possible values is the mechanism that allows the speed of the downstream front to select the same value as for the upstream front. As a result, puffs remain localized while travelling along the pipe. However, for r larger than the critical value, the upper-branch fixed point

no longer permits downstream fronts to occur at $u_f = u_c$, as seen in Fig. 2e. This restricts the possible values of s and, hence, the possible speeds of the downstream front to the bold portion of the branch illustrated in Extended Data Fig. 2a. Hence, as r passes through the critical point there is an abrupt change in the allowed values of the downstream front speeds, from an infinite to a finite range. Without nonlinear advection, the abrupt change is manifested as a discontinuous change in the speed of the downstream front. With nonlinear advection, there is still a discontinuous change to the allowed values of s, but the speeds are smaller than those of the upstream front, so the discontinuity in allowed solutions is masked.

Combining pipe and duct data. To combine data from pipe and duct flow into a single plot, it is necessary to determine specific Reynolds numbers and speeds from measured data (see Fig. 3), which will then be used to align the data from the two flows. Although the procedure is informed from the model analysis, it requires only measured data and the same procedure could be applied to data from other shear flows.

Extended Data Figure 3a, c shows data from pipe and duct flow, respectively, plotted with the upstream speeds reflected about the neutral speed, labelled C_0. The value of C_0 is determined to be that for which reflected upstream data coincides with the downstream data at sufficiently large Reynolds number. Extended Data Figure 3b, d shows the same data, but with model speeds (determined subsequently) also plotted as a visual aid. In the case of pipe flow, it is possible to determine C_0 to better that 2% accuracy. For duct flow, we estimate that the downstream front speed has not quite reached the reflected upstream speed at the highest Reynolds number accessible to present experiments. Nevertheless, C_0 is quite well determined. From the same plots, the value of the Reynolds number R_0 at which the upstream front obtains the neutral speed C_0, is easily determined.

We also determine the Reynolds number R_1 from the data, where the downstream weak front first deviates from the downstream front. This can in principle be determined solely from the data, but using model fits to the weak branch gives further confidence in the determined values. C_1 is the front speed at R_1.

Once the values (R_0, C_0) and (R_1, C_1) have been found for each flow, the data is collated by plotting each data set such that these two points collapse, as seen in Extended Data Fig. 4. This is equivalent to simply choosing the origin and scaling the axes for the two flows. The upstream and strong-downstream fronts each coincide, whereas the weak-front branch does not.

Determining model parameters. There are three model parameters, D, ζ and ϵ, to be determined to quantitatively relate the model speeds to the measured data for each flow.

The generic model cannot be expected to predict the flow-specific values R_0, R_1, C_0 and C_1, and, moreover, there is nothing universal about these values. Instead, given these flow-specific values, the model should capture the form of the various branches seen in the combined data of Extended Data Fig. 4. When fitting model parameters, it is useful to plot the combined data in terms of the reduced Reynolds number and reduced speed

$$\frac{R-R_0}{R_1-R_0}, \quad \frac{1}{2}\frac{C-C_0}{C_0-C_1} \quad (12)$$

which requires only relabelling of the axes in Extended Data Fig. 4 to shift the neutral speed C_0 to zero and scale the onset of the weak front to the point $(R_1, C_1) = (1, -1/2)$. As will become apparent, the reason for including $1/2$ in the reduced speed is so that model speeds are typically about half those of the reduced speeds from the experimental data.

We first consider the value of the parameter D. We select D so as to fix a simple relationship between model and measured quantities for both flows. Specifically, in Extended Data Fig. 5a, we plot the combined pipe and duct data together with the asymptotic results from the model for different values of D. The model results

are plotted directly in terms of the model Reynolds number r and the model speed shifted by the model neutral speed, $c - c_0$. For $D = 0.13$ the upstream front and strong branches match the combined experimental data extremely well. Note that the strong and weak asymptotic curves in Extended Data Fig. 5a are independent of the other two model parameters, ϵ and ζ.

Using only one parameter, D, and fixing its value to 0.13, the model not only fits the upstream and strong-downstream front speeds very well for both flows, but a simple relationship between model and experimental data are fixed, namely

$$r = \frac{R-R_0}{R_1-R_0}, \quad c-c_0 = \frac{1}{2}\frac{C-C_0}{C_0-C_1} \quad (13)$$

Given the flow-specific values R_0, R_1, C_0 and C_1, equation (13) is inverted to obtain the Reynolds number R and speed C from the model values r and c; this is how we map the model results to Reynolds number and speed in Fig. 3.

The remaining two model parameters dictate the behaviour of the downstream fronts as they transition from weakly expanding to strongly expanding. Here pipe and duct flows differ and so the values of the fitting parameters will necessarily be different for the two flows; see Extended Data Fig. 5b.

The value of ϵ dictates how quickly the system jumps from the weak to the strong branch. Large values give smoother transitions while smaller values give more abrupt transitions. The value of ζ dictates how long the system follows the weak branch before transitioning to the strong branch. Larger values, as for pipe flow, result in a delay in transition, whereas smaller values, as for the fit to duct flow, result in more immediate transition. We did not apply a formal procedure for determining ζ and ϵ for each of the flows. Rather they were determined simply by eye. In both cases it is quite easy to adjust ζ and ϵ so that the transition from weak to strong fronts follows the measured data.

Control. The model suggests that the fully turbulent state can be destabilized by removing the upper turbulent fixed point as depicted in Extended Data Fig. 6a. In the model, this is achieved by forcing the variable u, which corresponds to the state of the shear profile. The reduction of u by forcing corresponds to a blunting of the shear profile.

To demonstrate that the fully turbulent state can indeed be destabilized by removing the turbulent fixed point, as suggested by the model, we performed a direct numerical simulation of pipe flow for $R = 5,000$. Initially the forcing is not applied and the flow is fully turbulent. Starting at time $t = 175\, d/U$, a global body force is gradually switched on (fully applied by time $t = 200\, d/U$), which blunts the velocity profile to a more plug-like form (the same forcing is used as in ref. 18). As can be seen, turbulent intensity subsequently decreases, and eventually the fully turbulent flow destabilizes and degenerates into localized turbulent patches, similar to the natural ones (puffs) at lower Reynolds number (below about 2,300) in the absence of any additional force.

31. Willis, A. P. & Kerswell, R. R. Turbulent dynamics of pipe flow captured in a reduced model: puff relaminarization and localized 'edge' states. *J. Fluid Mech.* **619,** 213–233 (2009).
32. Wygnanski, I., Sokolov, M. & Friedman, D. On transition in a pipe. Part 2. The equilibrium puff. *J. Fluid Mech.* **69,** 283–304 (1975).
33. Holzner, M., Song, B., Avila, M. & Hof, B. Lagrangian approach to laminar–turbulent interfaces in transitional pipe flow. *J. Fluid Mech.* **723,** 140–162 (2013).
34. Kerswell, R. R. Recent progress in understanding the transition to turbulence in a pipe. *Nonlinearity* **18,** R17–R44 (2005).
35. Avila, M., Mellibovsky, F., Roland, N. & Hof, B. Streamwise-localized solutions at the onset of turbulence in pipe flow. *Phys. Rev. Lett.* **110,** 224502 (2013).
36. Chantry, M., Willis, A. P. & Kerswell, R. R. Genesis of streamwise-localized solutions from globally periodic traveling waves in pipe flow. *Phys. Rev. Lett.* **112,** 164501 (2014).

a No advection

b Nonlinear advection
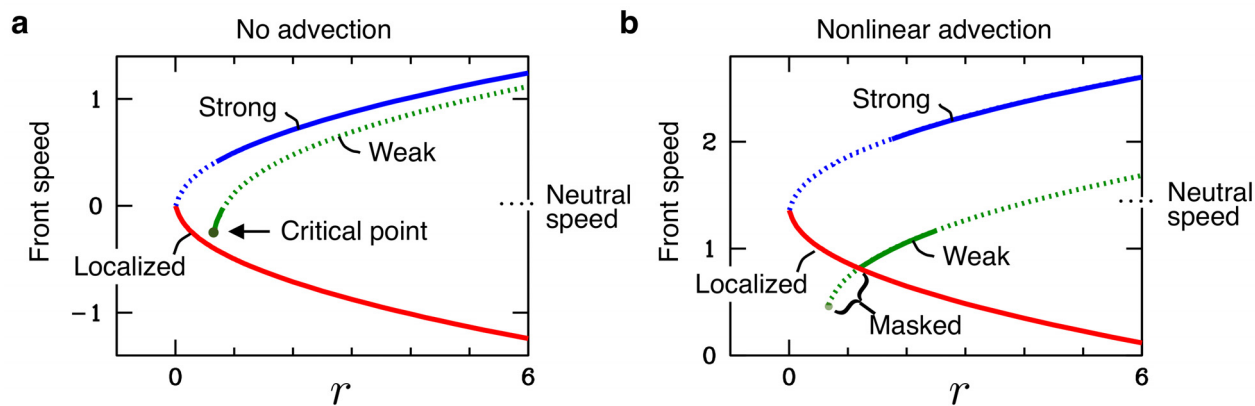
Extended Data Figure 1 | Speed of model fronts in the asymptotic limit of sharp fronts. a, b, Speeds as a function of model Reynolds number *r* both without (**a**) and with (**b**) advection. Although strong downstream fronts cannot exist and have no physical meaning below the formation of the upper branch fixed point, the expression for strong front speeds in equation (10) still gives the speed that such a strong downstream front would have; these speeds are shown dashed. The effect of nonlinear advection in **b** is to mask the nominal critical point for the onset of fully turbulent flow. The neutral speed is naturally displaced from the mean speed $U = 1$.
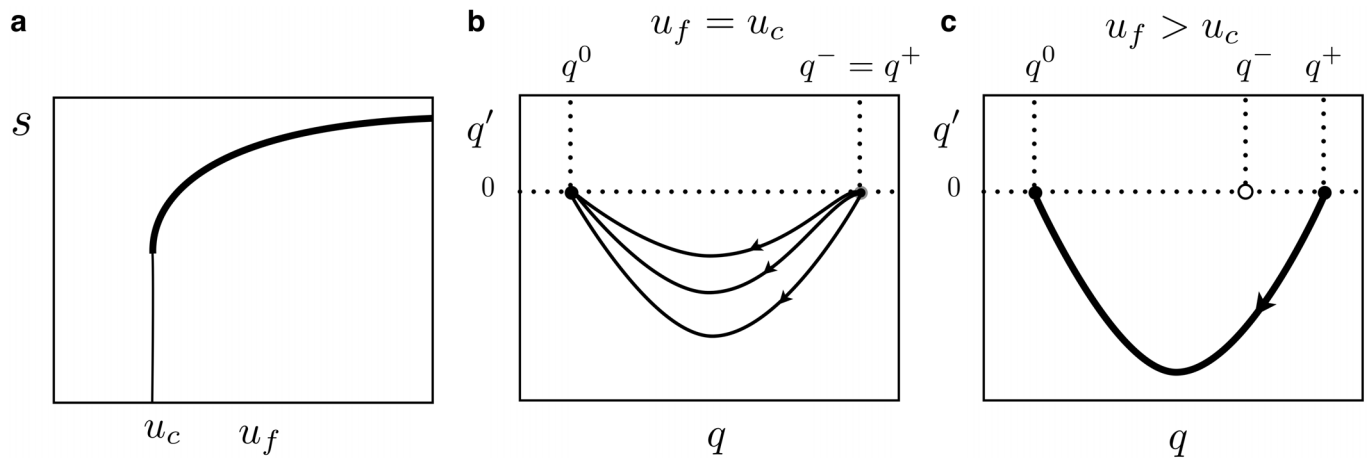
Extended Data Figure 2 | Front speeds at critical point. Sketch illustrating solutions to the boundary value problem in equation (8) for a downstream front near the critical point. **a**, Eigenvalue s as a function of u_f. u_c is the value of u_f such that $q^- = q^+$. For this value there are infinitely many possible eigenvalues s, indicated by the thin line. **b**, **c**, Phase planes (q, q') showing solutions for the second order differential equation (8). Downstream fronts are heteroclinic connections from the upper fixed point q^+ to the lower fixed point q^0. When $u_f = u_c$ and hence $q^- = q^+$, the upper fixed point is not hyperbolic and there are infinitely many connections, each corresponding to a value of s. When $u > u_c$, q^+ is hyperbolic and there is a unique connection and hence a unique value of s.
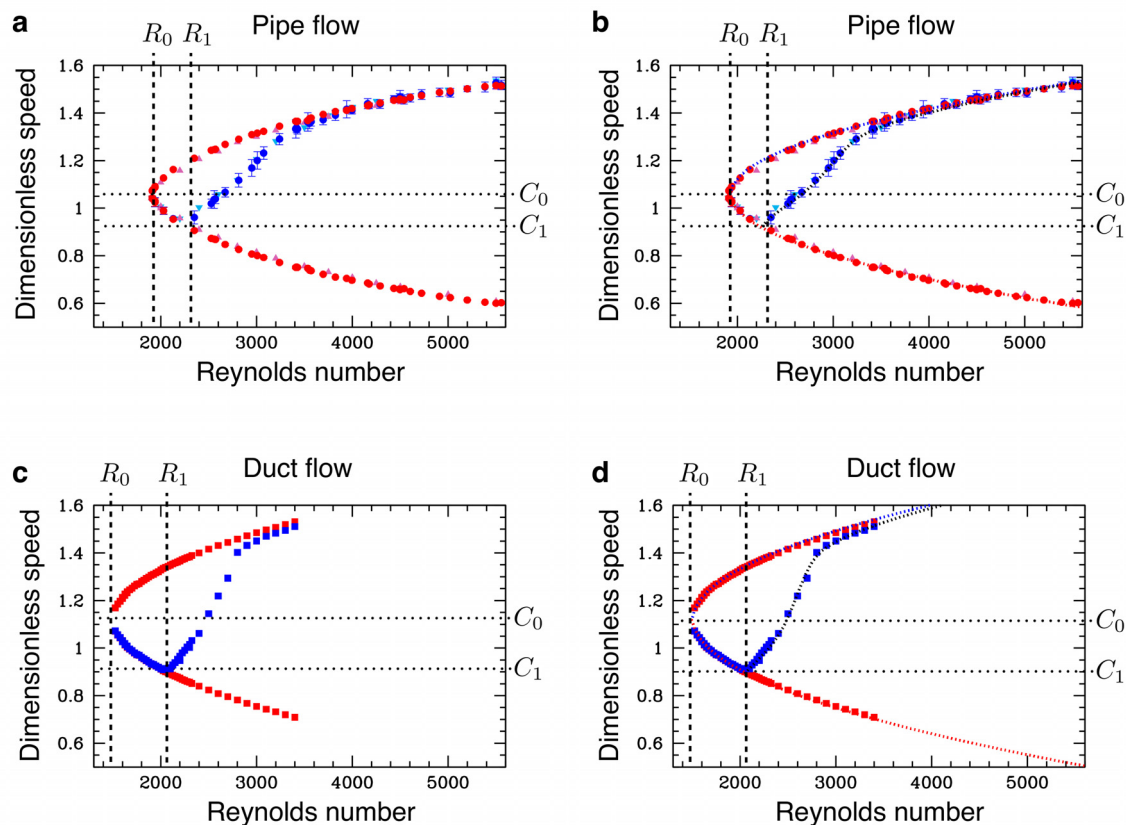
Extended Data Figure 3 | Determination of corresponding Reynolds numbers and speeds for pipe and duct flow. a–d, Speeds from pipe (**a**, **b**) and duct flow (**c**, **d**) are plotted, as in Fig. 1c, but additionally with the upstream front speeds reflected about the neutral speed C_0. **a**, **c**, Experimental and simulation data only; **b**, **d**, model fits to the experimental and simulation data. The determined values for R_0, R_1, C_0 and C_1 are: $R_0 = 1,920$, $C_0 = 1.06$, $R_1 = 2,250$ and $C_1 = 0.92$ for pipe flow, and $R_0 = 1,490$, $C_0 = 1.12$, $R_1 = 2,030$ and $C_1 = 0.90$ for duct flow.
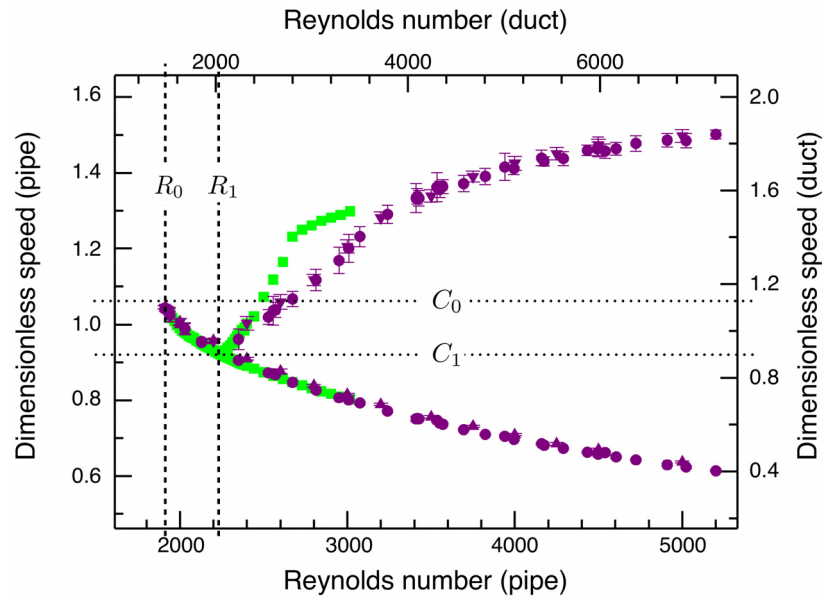
Extended Data Figure 4 | Combining pipe and duct data. Pipe and duct flow are plotted together using different axes. The data are plotted so that the two points (R_0, C_0) and (R_1, C_1) align for each data set; for example, $(R_1, C_1) = (2{,}250, 0.92)$ for pipe flow is aligned with $(R_1, C_1) = (2{,}030, 0.90)$ for duct flow, bringing into alignment the onset of weak fronts.

a Parameter D
pipe and duct flow

b Parameters ζ, ϵ

pipe flow
$\zeta = 0.79, \epsilon = 0.05, 0.1, 0.2, 0.4$

duct flow
$\zeta = 0.56, \epsilon = 0.05, 0.1, 0.2, 0.4$
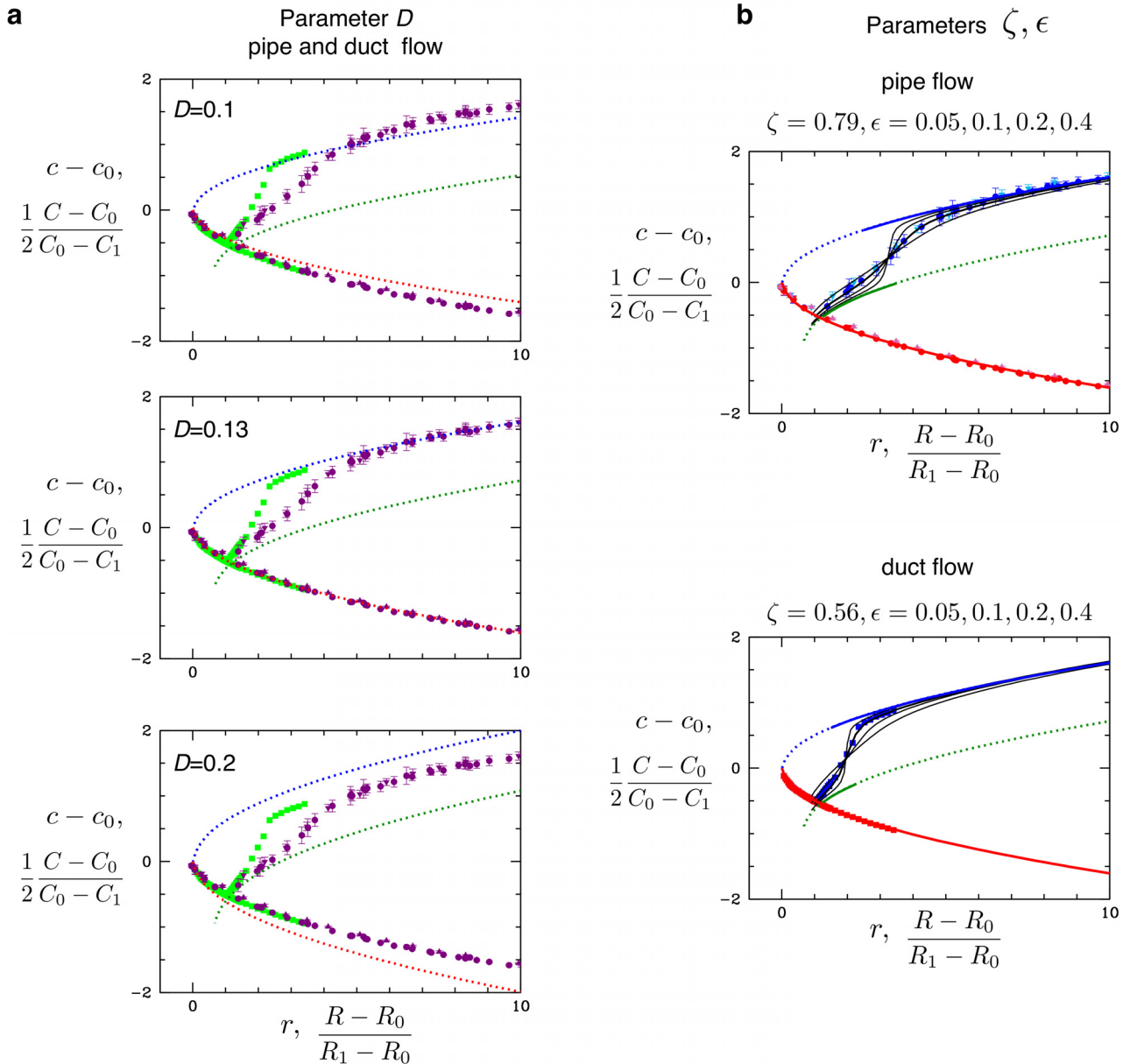
Extended Data Figure 5 | Determination of model parameters for pipe and duct flow. a, Determination of D. Points are data from pipe and duct flow (as in Extended Data Fig. 4) here plotted in terms of reduced Reynolds number $(R - R_0)/(R_1 - R_0)$ and reduced speed $(C - C_0)/[2(C_0 - C_1)]$. Dashed curves are asymptotic speed curves (as in Extended Data Fig. 1) plotted in terms of model Reynolds number r and speed $c - c_0$. For $D = 0.13$ there is very good

agreement between the data and the model. This choice of D fixes the asymptotic branches (dashed curves). **b**, Determination of ζ and ϵ. Pipe and duct flow are necessarily considered separately. In each case, downstream branches are shown for four values of ϵ. Smaller values yield more abrupt transitions between weak and strong branches.

a

b
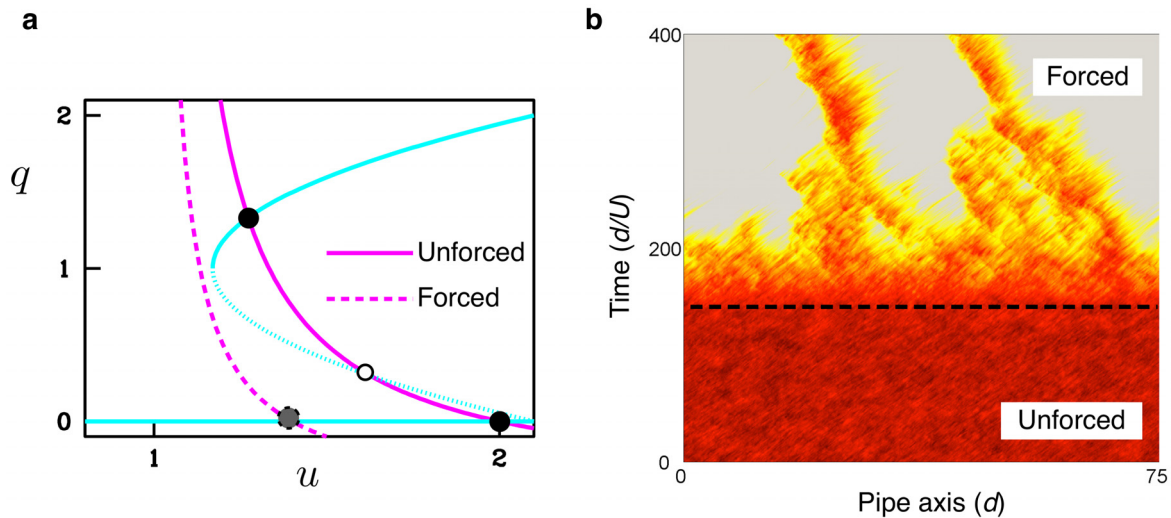
Extended Data Figure 6 | Illustration of control by removing the turbulent fixed point. a, Control concept illustrated in the model phase plane. Without forcing (that is, without control), there is an upper-branch fixed point (upper intersection of nullclines) corresponding to fully turbulent flow. Applying an additive forcing term to the u equation corresponds to forcing the shear profile and blunting its shape. This can remove the turbulent fixed point thus eliminating fully turbulent flow. **b**, Proof of concept in a direct numerical simulation of pipe flow at $R = 5,000$. Without forcing the flow is fully turbulent. A global body force is applied that blunts the velocity profile to a more plug-like form. Subsequently, only localized turbulent patches remain, reminiscent of those at much lower R.

Extended Data Table 1 | The domain size and resolution for the simulations at all the Reynolds numbers we considered

R	$L_x(d)$	N	K	M	R	$L_x(d)$	N	K	M
1920	24π	48	640	32	3200	180	72	2560	54
2000	24π	48	768	40	3500	180	72	2560	54
2200	133	48	768	40	3750	133	72	2048	54
2300	133	64	1536	40	4000	133	72	2048	54
2400	133	64	2048	48	4500	180	80	3072	64
2800	180	72	2560	48	5000	180	80	3072	64
2600	133	64	2048	48	5500	180	96	3840	80

In physical space there are $3K$ and $3M$ grid points in axial and azimuthal directions, respectively. N is the number of grid points across the pipe radius $d/2$.

LETTER

Observation of non–Hermitian degeneracies in a chaotic exciton–polariton billiard

T. Gao[1], E. Estrecho[1], K. Y. Bliokh[1,2], T. C. H. Liew[3], M. D. Fraser[2], S. Brodbeck[4], M. Kamp[4], C. Schneider[4], S. Höfling[4,5], Y. Yamamoto[6,7], F. Nori[2,8], Y. S. Kivshar[1], A. G. Truscott[1], R. G. Dall[1] & E. A. Ostrovskaya[1]

Exciton-polaritons are hybrid light–matter quasiparticles formed by strongly interacting photons and excitons (electron–hole pairs) in semiconductor microcavities[1–3]. They have emerged as a robust solid-state platform for next-generation optoelectronic applications as well as for fundamental studies of quantum many-body physics. Importantly, exciton-polaritons are a profoundly open (that is, non-Hermitian[4,5]) quantum system, which requires constant pumping of energy and continuously decays, releasing coherent radiation[6]. Thus, the exciton-polaritons always exist in a balanced potential landscape of gain and loss. However, the inherent non-Hermitian nature of this potential has so far been largely ignored in exciton-polariton physics. Here we demonstrate that non-Hermiticity dramatically modifies the structure of modes and spectral degeneracies in exciton-polariton systems, and, therefore, will affect their quantum transport, localization and dynamical properties[7–9]. Using a spatially structured optical pump[10–12], we create a chaotic exciton-polariton billiard—a two-dimensional area enclosed by a curved potential barrier. Eigenmodes of this billiard exhibit multiple non-Hermitian spectral degeneracies, known as exceptional points[13,14]. Such points can cause remarkable wave phenomena, such as unidirectional transport[15], anomalous lasing/absorption[16,17] and chiral modes[18]. By varying parameters of the billiard, we observe crossing and anti-crossing of energy levels and reveal the non-trivial topological modal structure exclusive to non-Hermitian systems[9,13–22]. We also observe mode switching and a topological Berry phase for a parameter loop encircling the exceptional point[23,24]. Our findings pave the way to studies of non-Hermitian quantum dynamics of exciton-polaritons, which may uncover novel operating principles for polariton-based devices.

Studies of open quantum systems go back to Gamow's theory of nuclear α-decay developed in the early days of quantum mechanics[4]. Indeed, metastable states of a single quantum particle in a spherically symmetric potential well with semi-transparent barriers decay in time, and therefore are characterized by complex energies. Furthermore, introducing a 2D potential well with non-trivial geometry, that is, a quantum billiard, results in strongly correlated energy levels and transition to quantum chaos[7,19,21,25–28]. Spectral degeneracies crucially determine transport and dynamical properties in both non-Hermitian and chaotic wave systems[7–9,15–17]. In chaotic and disordered wave systems, spectral degeneracies underpin statistical properties and quantum phase transitions from localized to delocalized dynamics[8,9]. In non-Hermitian (including PT-symmetric) systems, non-trivial topology of eigenmodes and unusual transport properties in the vicinity of exceptional points[15–17,19] are currently under investigation. Basic non-Hermitian or stochastic dynamics have so far been studied in the context of microwave[9,18–20,24], optical[15–17,19,21], atomic[22,26,27] and electron[25,28] waves. However, the concepts of non-Hermiticity and quantum chaos remain largely separated from each other, owing to the lack of a simple

quantum system in which both features would be readily accessible. Moreover, it is challenging to produce artificial complex potentials with gain and loss for classical waves, as well as to observe nanoscopic electron states in solids.

Microcavity exciton-polaritons represent a unique quantum macroscopic system, which combines the main advantages of light and matter waves[1–3]. Being bosons, exciton-polaritons can display collective quantum behaviour, Bose–Einstein condensation (BEC), when they occupy a single-particle quantum state in massive numbers. Exciton-polaritons have provided a very accessible system for studies of collective quantum behaviour because they condense at temperatures ranging from 10 K to room temperature (compared to nanokelvins for neutral atoms) and do not require painstaking isolation from the environment.

The schematics of exciton-polariton condensation under continuous-wave incoherent optical excitation conditions[1] are shown in Fig. 1a. The optical pump, far detuned from the exciton resonance in the cavity, effectively creates an incoherent reservoir of 'hot', exciton-like polaritons. Above a threshold density of the reservoir, relaxation and stimulated scattering into the coherent BEC state of exciton-polaritons dominate the dynamics. The continuously pumped condensate decays and releases coherent photons, which escape the cavity carrying all information about the condensed state. The interactions between the reservoir and condensed exciton-polaritons are responsible for the formation of effective pump-induced potentials[10–12]. Thus, the macroscopic matter wavefunction is shaped by an optical pump and spatially resolved via free-space optical microscopy. This enables us to clearly observe and control non-Hermitian and irregular quantum dynamics.

We use a structured optical pump[10–12] to create a non-Hermitian potential in the shape of a Sinai billiard[7] with a circular defect of radius R (see Fig. 1b) for condensed exciton-polaritons (see Methods for details). In our experiment, the billiard has 'soft' (inelastic) walls of a finite width and height. The main properties of eigenstates of the exciton-polariton condensate in the billiard can be described by a linear Schrödinger equation with a complex two-dimensional potential $V(\boldsymbol{r}) = V'(\boldsymbol{r}) + iV''(\boldsymbol{r})$. Here the real part of the potential, $V'(\boldsymbol{r}) \propto P(\boldsymbol{r})$, is the potential barrier shaped as a Sinai billiard boundary with a Gaussian envelope. The optical pump rate, $P(\boldsymbol{r})$, is induced by the strong repulsive interaction between the excitonic reservoir populated by the pump and the polariton BEC[10–12]. The imaginary part of the potential, $V''(\boldsymbol{r}) \propto P(\boldsymbol{r}) - \gamma$, combines the gain profile produced by the same optical pump $P(\boldsymbol{r})$ with the spatially uniform loss γ due to polariton decay (Fig. 1b). Despite the strong polariton–polariton interactions, the corresponding nonlinearity mostly affects the relative population of the energy eigenstates, as well as the overall blueshift (see Methods).

Changing the radius of the defect, R, varies the geometry of the billiard and hence affects the energy levels. Figures 1c and d show the

[1]Research School of Physics and Engineering, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. [2]Center for Emergent Matter Science, RIKEN, Wako-shi, Saitama 351-0198, Japan. [3]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore. [4]Technische Physik and Wilhelm-Conrad-Röntgen Research Center for Complex Material Systems, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany. [5]SUPA, School of Physics and Astronomy, University of St Andrews, St Andrews KY16 9SS, UK. [6]ImPACT Project, Japan Science and Technology Agency, Chiyoda-ku, Tokyo 102-0076, Japan. [7]Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305-4085, USA. [8]Physics Department, University of Michigan, Ann Arbor, Michigan 48109-1040, USA.
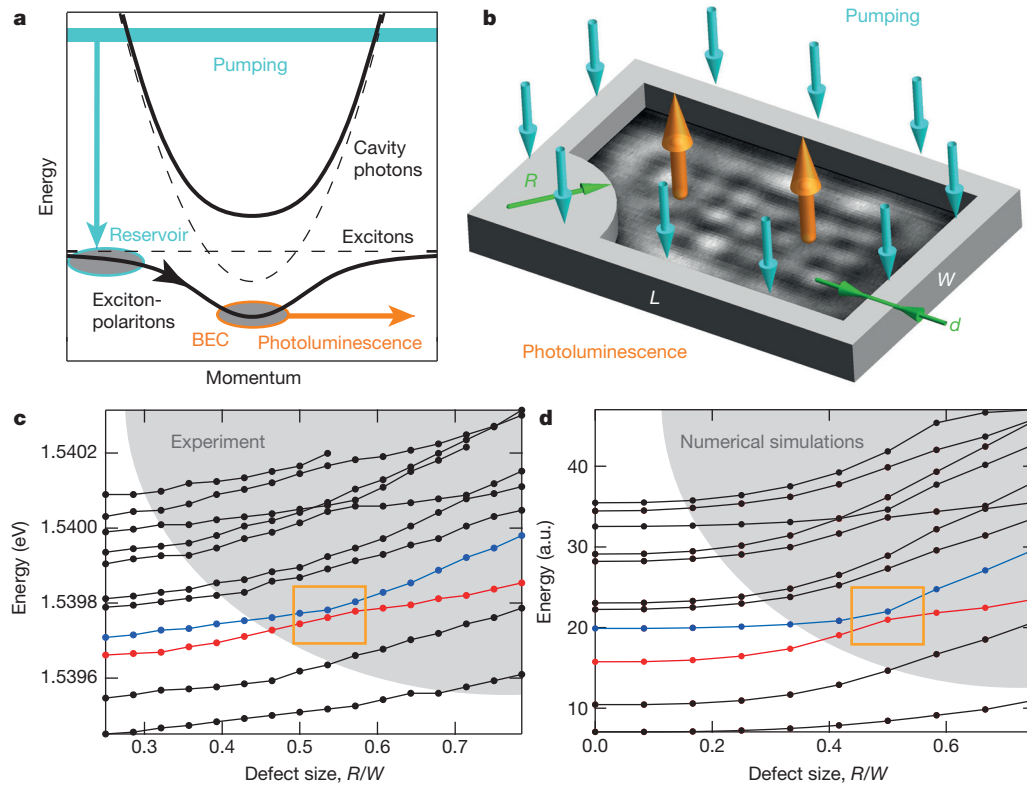
Figure 1 | Non-Hermitian exciton-polariton Sinai billiard and its spectrum. **a**, Exciton-polariton dispersion showing the upper and lower branches (solid lines) formed owing to hybridization of the cavity photon and exciton modes (dashed lines). The incoherent excitonic reservoir is continuously replenished by the optical pump (represented by the cyan arrow) and 'feeds' the polariton BEC (black arrow). The polariton BEC decays into cavity photoluminescence (orange arrow)[1–3]. **b**, Schematics of the exciton-polariton Sinai billiard formed in the plane of a quantum well embedded into the microcavity (see Methods). The barrier is induced by the optical pump via the excitonic reservoir, and the square modulus of the wavefunction of the confined polariton BEC (shown in greyscale inside the billiard) is imaged via the photoluminescence. The billiard dimensions are fixed as $W = 14\,\mu m$, $L = 23\,\mu m$, the radius of the defect R is varied from 0 to W, and the thickness of the walls d is varied from $3\,\mu m$ to $7\,\mu m$ (see Methods). **c**, **d**, Experimentally measured (**c**) and numerically simulated (**d**) spectra $E(R/W)$ for the first 11 modes of the billiard in arbitrary units (a.u.). With growing R, numerous degeneracies and quasi-degeneracies proliferate in the grey area, which is a signature of the transition to quantum chaos in the Hermitian Sinai billiard[7]. Topological properties of two near-degenerate modes (red and blue in the orange rectangles) are analysed in detail in Figs 2–4.

experimentally measured and numerically computed energy spectra $E(R)$ of the first 11 levels as a function of R. Variations of the shape of the 2D potential tune eigenvalues of different modes at different rates, and as a result some energy levels approach each other at certain values of R. It can be seen (Fig. 1c, d) that multiple degeneracies (or near-degeneracies) appear in the spectrum. In a 'hard-wall' Hermitian Sinai billiard, the proliferation of degeneracies is a signature of the transition from regular to chaotic dynamics[7]. Although our exciton-polariton billiard has 'soft' walls and can generically exhibit mixed regular-chaotic behaviour[27], we clearly observe multiple degeneracies similar to the 'hard-wall' case[7]. In Hermitian billiards, the levels generically avoid crossing (that is, they anti-cross) in the vicinity of degeneracies, which correspond to the average level repulsion and Wigner distribution of the nearest-neighbour energy spacings[8]. In contrast, the non-Hermitian systems can exhibit both crossings and anti-crossings of levels[9,19–22]. This is because the energy eigenvalues in non-Hermitian systems are complex: the real part and imaginary parts correspond to the real energies and linewidths of the modes, respectively. A crossing of the energies is accompanied by an anti-crossing of the linewidths and vice versa. In our experiment, we measure the spectral profile of the cavity photoluminescence at a particular spatial position and extract both peak energies and widths of spectral resonances (see Methods). Crossings as well as anti-crossings of real energy levels are clearly seen both in experiments (Fig. 1c) and numerical simulations (Fig. 1d).

To observe the transition between crossing and anti-crossing for the same near-degenerate pair of eigenvalues, a second control parameter

needs to be varied. In our exciton-polariton billiard, this additional parameter is the thickness, d, of the billiard walls. Provided the internal area of the billiard remains unchanged, this parameter does not affect the geometry of the billiard and primarily controls the imaginary part V'' of the non-Hermitian potential barrier. Figure 2 shows one pair of billiard modes highlighted in Fig. 1c in the vicinity of a near-degeneracy for two values of the control parameter d. One can clearly see the anti-crossing (crossing) behaviour of the real (imaginary) parts of the complex eigenenergies in the billiard with thick walls (Fig. 2a, c) and the opposite behaviour for the thin-wall billiard (Fig. 2b, d).

Importantly, the energy-resolved real-space imaging of the photoluminescence provides all the information about complex eigenvalues as well as the spatial structure of the eigenmodes (wavefunctions). In particular, the levels shown in Fig. 2 correspond (at $R = 0$) to the third mode with three horizontal lobes and the fourth mode with two vertical lobes. The experimentally imaged and calculated spatial profiles of these eigenmodes are shown as insets in Fig. 2a, b along the eigenenergy curves. We observe that the two modes are hybridized and therefore change their spatial profiles in the near-degeneracy region, and 'exchange' their spatial profiles after passing it.

The behaviour of two billiard modes in the vicinity of a degeneracy can be described by a simple model of a two-level system with an effective coupling (see Methods). The corresponding non-Hermitian Hamiltonian reads[9,18–22]:

$$\hat{H} = \begin{pmatrix} \tilde{E}_1 & q \\ q^* & \tilde{E}_2 \end{pmatrix}, \quad \tilde{E}_{1,2} = E_{1,2} - i\Gamma_{1,2} \qquad (1)$$
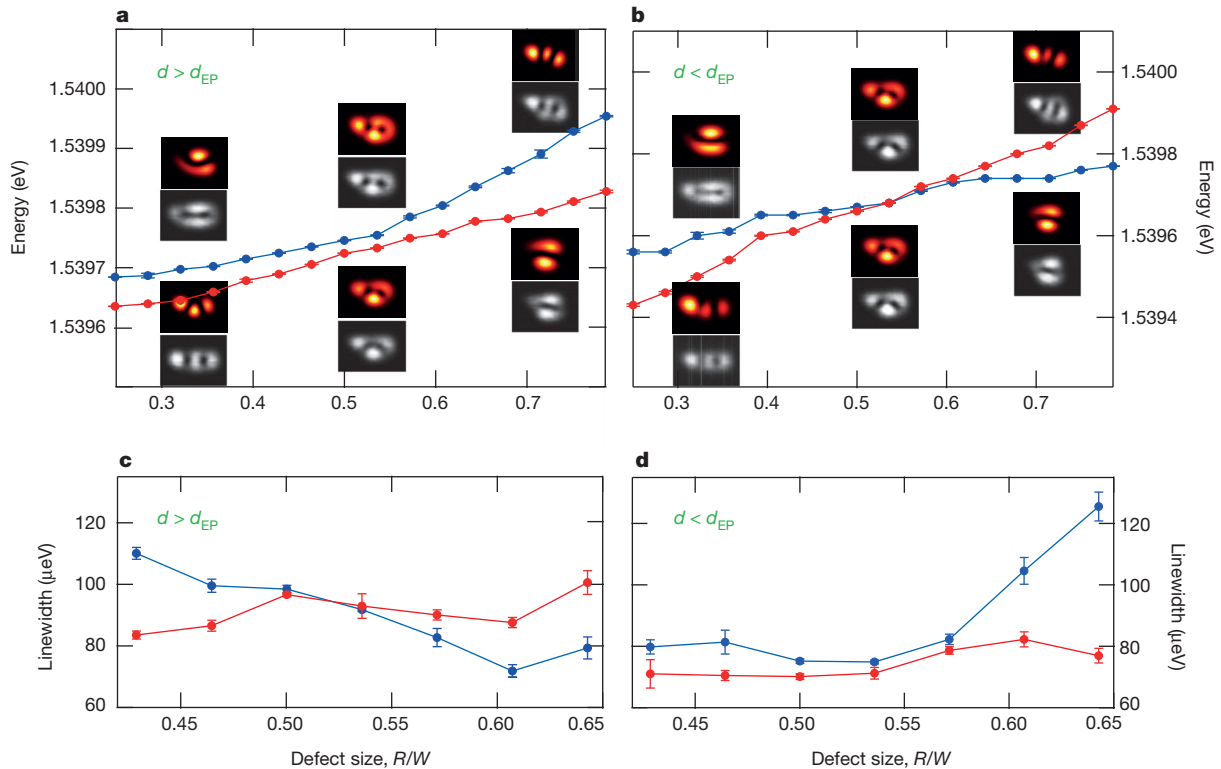
Figure 2 | Crossing and anti-crossing for two near-degenerate modes. These modes are boxed in Fig. 1c, d. **a–d**, Experimentally observed anti-crossing (**a**) and crossing (**b**) of eigenenergies of two modes in the spectrum of the exciton-polariton Sinai billiard with varying parameter R (see Fig. 1) for thick, $d \approx 6\,\mu m$ (**a**, **c**), and thin, $d \approx 4\,\mu m$ (**b**, **d**), billiard walls; d_{EP} is the value corresponding to the exceptional point. Panels **c** and **d** show the corresponding crossing and anti-crossing of the linewidths (that is, imaginary parts of the complex eigenvalues). The error bars in **a–d** originate from numerical fitting of the spectroscopic data (see Methods). The upper (lower) inset panels in **a** and **b** illustrate the numerically calculated (experimentally imaged) spatial structure of the eigenmodes at different values of the parameter R. Details of the hybridization region are given in Methods.

Here $\tilde{E}_{1,2}$ are the complex eigenvalues of two uncoupled modes (with $E_{1,2}$ being the real energies and $\Gamma_{1,2}$ being the decay/gain rates), whereas q characterizes the coupling between these two modes (the star stands for complex conjugation). We will also use the mean complex energy $\tilde{E} = (\tilde{E}_1 + \tilde{E}_2)/2 \equiv E - i\Gamma$, and the complex energy difference $\delta\tilde{E} = (\tilde{E}_2 - \tilde{E}_1)/2 \equiv \delta E - i\delta\Gamma$. The eigenvalues of the Hamiltonian (equation (1)) are $\lambda_{1,2} = \tilde{E} \pm \sqrt{\delta\tilde{E}^2 + |q|^2}$; their real and imaginary parts, which depend on the parameters $\delta\tilde{E} = (\delta E, \delta\Gamma)$, are shown in Fig. 3. These complex eigenvalues coalesce, $\lambda_1 = \lambda_2$, at the exceptional points (EPs)[13–22], where $i\delta\tilde{E}_{EP} = \pm|q|$. At these points, the eigenstates also coalesce and form a single chiral mode[13,14,18]. Assuming that the coupling constant q is fixed, the exceptional points appear in the parameter plane as $(\delta E_{EP}, \delta\Gamma_{EP}) = (0, \pm|q|)$. We assume $\delta\Gamma > 0$ in our range of parameters, so that there is only one exceptional point in the domain of interest. The exceptional point
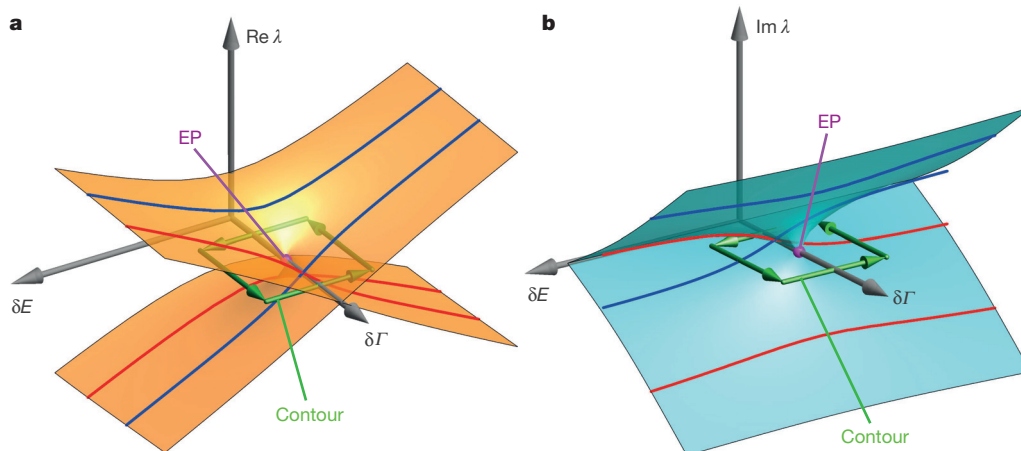
Figure 3 | Eigenvalues of a two-level non-Hermitian model in the vicinity of the exceptional point. a, b, Real (**a**) and imaginary (**b**) parts of the eigenvalues $\lambda_{1,2}$ of the model (equation (1)) as functions of two parameters, δE and $\delta\Gamma$. The exceptional point (EP) is shown in magenta. The crossing and anti-crossing of the real and imaginary parts of the eigenvalues as functions of δE, for $\delta\Gamma < \delta\Gamma_{EP}$ and $\delta\Gamma > \delta\Gamma_{EP}$, are shown in red and blue. This is in correspondence with the experimentally observed behaviour in Fig. 2. Traversing along the green contour encircling the exceptional point in the $(\delta E, \delta\Gamma)$ plane reveals the non-trivial topology of eigenmodes, as shown in Fig. 4.
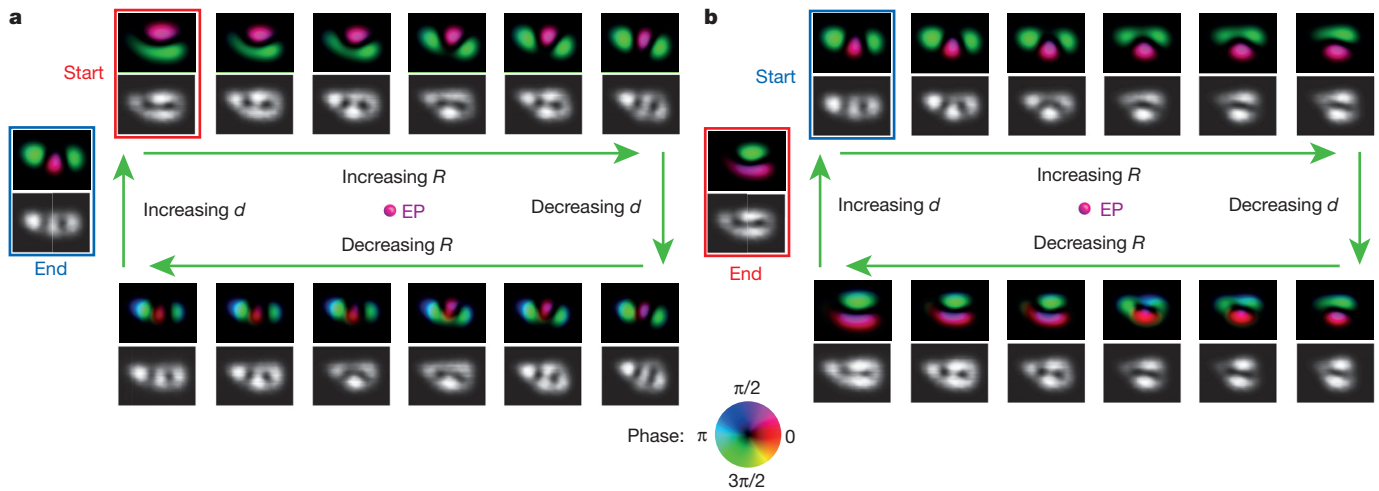
Figure 4 | Observation of the topological Berry phase acquired after circling around the exceptional point in the parameter plane. Transmutations of spatial distributions (black-and-white panels) of the selected eigenmode (from the pair shown in Fig. 2) along the closed contour in the parameter space $(R, d) \sim (\delta E, \delta \Gamma)$ encircling the exceptional point (see Fig. 3). Parameters are not varied in time during the measurements, and each distribution corresponds to the stationary mode at the corresponding parameter values.

a, b, The first loop (**a**) shows the transition to a different branch (mode) through the hybridization region (see explanations in text); the second loop (**b**) returns the mode to the original one with a π topological phase shift[23,24]. The phases (colour panels) are inferred from comparison with the numerically calculated modes. The modes corresponding to the 'start' and 'end' points of the loop on the red (blue) branch in Figs 2a, b and 3a are boxed in red (blue).

can be encircled in the $(\delta E, \delta \Gamma)$ plane by varying these two parameters, as seen in Fig. 3.

Two parameters of the model, $(\delta E, \delta \Gamma)$, approximately correspond to the varying parameters (R, d) of our exciton-polariton billiard. The radius R mostly affects the real part of the potential, $V'(\mathbf{r})$, and hence the energy difference between the modes. Increasing R corresponds to a tighter spatial confinement and therefore to increasing δE. In turn, the thickness d of the billiard walls controls the gain/loss profile $V''(\mathbf{r})$. Different modes have different spatial overlaps with the imaginary potential $V''(\mathbf{r})$, and, therefore, are characterized by different integral (spatially averaged) dissipation parameters $\Gamma_{1,2}$ (see Methods). In our case, increasing d corresponds to decreasing $\delta \Gamma$. The effective coupling q in our model (equation (1)) is determined by the spatial overlap between the two modes away from the hybridization region[9]. The red and blue curves in Fig. 3 show the crossing/anti-crossing behaviour of the real and imaginary parts of the eigenvalues versus the energy difference δE for two values of the dissipation parameter: $\delta \Gamma < \delta \Gamma_{EP}$ and $\delta \Gamma > \delta \Gamma_{EP}$. This behaviour is perfectly consistent with that in the experimental Fig. 2, which means that our range of varying parameters includes the exceptional point.

The structure of the complex eigenvalues in the vicinity of the exceptional point reveals non-trivial topology of a branch-point type[13–22], shown in Fig. 3. Therefore, continuous encircling of the non-Hermitian degeneracy in the two-parameter plane (for example, along the green contour in Fig. 3) results in the transition to the other branch. When the contour is traversed twice, we return to the original mode, most significantly with a topological phase shift of π. This phase shift is the manifestation of the Berry phase resulting from encircling of a non-Hermitian degeneracy in a two-dimensional parameter space[23,24]. We use the method suggested in the microwave experiment[24] to trace the above topological structure of two modes in the vicinity of the exceptional point. We compare the eigenmodes at neighbouring values of parameters $(\delta E, \delta \Gamma) \sim (R, d)$ along the contour encircling the exceptional point (see Fig. 3). Notably, we do not consider adiabatic evolution of modes due to variations of the parameters (R, d) in time; such evolution would be accompanied by unavoidable non-adiabatic transitions in the non-Hermitian case[29,30]. Rather, we examine the natural topological structure and geometrical connection of stationary modes depending on the parameter values.

Figure 4 depicts the experimentally measured intensities and the corresponding numerically simulated phase profiles of the two modes from Fig. 2 for the parameter values lying on the contour encircling the exceptional point (Fig. 3). In Fig. 4a, we start on the upper branch (blue in Figs 2a and 3a) at $R < R_{EP}$, $d > d_{EP}$ and trace the eigenmode transmutation as the radius is increased to $R > R_{EP}$. This takes us from the vertical two-lobe mode, through the anti-crossing, to the horizontal three-lobe mode (still on the blue upper branch). Then, we decrease the thickness to $d < d_{EP}$ and stay on the same horizontal three-lobe mode, which now corresponds to the red branch in Figs 2b and 3a. Next, reducing the radius R takes this mode through the crossing and recovers its three-lobe structure. Increasing d closes the loop. Thus, the continuous transformation brought us from the vertical two-lobe mode ('start' in Fig. 4a) to the horizontal three-lobe mode ('end' in Fig. 4a) at the same values of the parameters. Repeating this traverse one more time (Fig. 4b) returns us to the original vertical two-lobe mode, but now with the π topological phase shift (clearly seen in the simulated phase profiles). The experimental density distribution of the modes is in very good agreement with that calculated numerically. Therefore we can associate the phase structure of the simulated spatial modes with the experimental mode profiles[24].

Thus, we have demonstrated the creation of highly controllable complex (non-Hermitian) potentials for exciton-polaritons, and implemented a chaotic non-Hermitian exciton-polariton billiard with multiple spectral degeneracies. We have provided detailed experimental observations of the non-trivial behaviour of complex eigenvalues and eigenmodes in the vicinity of an exceptional point. These include crossing/anti-crossing transitions as well as mode switching and topological Berry phase when encircling the exceptional point in the two-parameter plane. Our results show that the inherent non-Hermitian nature of exciton-polaritons determines their basic properties, which are crucial for transport and quantum information processing. Therefore, these features should be taken into account in future studies and applications involving confinement and manipulation of exciton-polaritons. Most importantly, this complex quantum dynamics can bring novel functionality to polariton-based devices operating at the interface between photonics and electronics. Generally, exciton-polaritons offer a novel macroscopic quantum platform for studies of non-Hermitian physics and quantum chaos at the confluence of light and matter.

1. Kasprzak, J. *et al.* Bose–Einstein condensation of exciton polaritons. *Nature* **443,** 409–414 (2006).
2. Deng, H., Haug, H. & Yamamoto, Y. Exciton-polariton Bose–Einstein condensation. *Rev. Mod. Phys.* **82,** 1489–1537 (2010).
3. Carusotto, I. & Ciuti, C. Quantum fluids of light. *Rev. Mod. Phys.* **85,** 299–366 (2013).
4. Moiseyev, N. *Non-Hermitian Quantum Mechanics* (Cambridge Univ. Press, 2011).
5. Bender, C. M. Making sense of non-Hermitian Hamiltonians. *Rep. Prog. Phys.* **70,** 947–1018 (2007).
6. Savvidis, P. G. *et al.* Off-branch polaritons and multiple scattering in semiconductor microcavities. *Phys. Rev. B* **64,** 075311 (2001).
7. Berry, M. V. Quantizing a classically ergodic system: Sinai's billiard and the KKR method. *Ann. Phys.* **131,** 163–216 (1981).
8. Guhr, T., Müller-Groeling, A. & Weidenmüller, H. A. Random-matrix theories in quantum physics: common concepts. *Phys. Rep.* **299,** 189–425 (1998).
9. Bliokh, K. Y., Bliokh, Y. P., Freilikher, V., Genack, A. Z. & Sebbah, P. Coupling and level repulsion in the localized regime: from isolated to quasiextended modes. *Phys. Rev. Lett.* **101,** 133901 (2008).
10. Tosi, G. *et al.* Sculpting oscillators with light within a nonlinear quantum fluid. *Nature Phys.* **8,** 190–194 (2012).
11. Sanvitto, D. *et al.* All-optical control of the quantum flow of a polariton condensate. *Nature Photon.* **5,** 610–614 (2011).
12. Dall, R. *et al.* Creation of orbital angular momentum states with chiral polaritonic lenses. *Phys. Rev. Lett.* **113,** 200404 (2014).
13. Berry, M. V. Physics of non-Hermitian degeneracies. *Czech. J. Phys.* **54,** 1039–1047 (2004).
14. Heiss, W. D. The physics of exceptional points. *J. Phys. Math. Gen.* **45,** 444016 (2012).
15. Guo, A. *et al.* Observation of PT-symmetry breaking in complex optical potentials. *Phys. Rev. Lett.* **103,** 093902 (2009).
16. Peng, B. *et al.* Loss-induced suppression and revival of lasing. *Science* **346,** 328–332 (2014).
17. Sun, Y., Tan, W., Li, H.-Q., Li, J. & Chen, H. Experimental demonstration of a coherent perfect absorber with PT phase transition. *Phys. Rev. Lett.* **112,** 143903 (2014).
18. Dembowski, C. *et al.* Observation of a chiral state in a microwave cavity. *Phys. Rev. Lett.* **90,** 034101 (2003).
19. Cao, H. & Wiersig, J. Dielectric microcavities: model systems for wave chaos and non-Hermitian physics. *Rev. Mod. Phys.* **87,** 61–111 (2015).
20. Dembowski, C. *et al.* Experimental observation of the topological structure of exceptional points. *Phys. Rev. Lett.* **86,** 787–790 (2001).
21. Lee, S.-B. *et al.* Observation of an exceptional point in a chaotic optical microcavity. *Phys. Rev. Lett.* **103,** 134101 (2009).
22. Choi, Y. *et al.* Quasi-eigenstate coalescence in an atomic-cavity quantum composite. *Phys. Rev. Lett.* **104,** 153601 (2010).
23. Heiss, W. D. Phases of wave functions and level repulsion. *Eur. Phys. J. D* **7,** 1–4 (1999).
24. Dembowski, C. *et al.* Encircling an exceptional point. *Phys. Rev. E* **69,** 056216 (2004).
25. Akis, R., Ferry, D. K. & Bird, J. P. Wave function scarring effects in open stadium shaped quantum dots. *Phys. Rev. Lett.* **79,** 123–126 (1997).
26. Milner, V., Hanssen, J. L., Campbell, W. C. & Raizen, M. G. Optical billiards for atoms. *Phys. Rev. Lett.* **86,** 1514–1517 (2001).
27. Kaplan, A., Friedman, N., Anderson, M. & Davidson, N. Observation of islands of stability in soft wall atom-optics billiards. *Phys. Rev. Lett.* **87,** 274101 (2001).
28. Ponomarenko, L. A. *et al.* Chaotic Dirac billiard in graphene quantum dots. *Science* **320,** 356–358 (2008).
29. Uzdin, R., Mailybaev, A. & Moiseyev, N. On the observability and asymmetry of adiabatic state flips generated by exceptional points. *J. Phys. A* **44,** 435302 (2011).
30. Berry, M. V. & Uzdin, R. Slow non-Hermitian cycling: exact solutions and the Stokes phenomenon. *J. Phys. A* **44,** 435303 (2011).

Author Contributions E.A.O., T.G., E.E. and K.Y.B. conceived the idea for this research; T.G., E.E., M.D.F., R.G.D. and A.G.T. designed and built the experiment with conceptual contributions from E.A.O.; T.G., E.E. and R.G.D. collected and analysed experimental data; K.Y.B, E.E., T.C.H.L. and E.A.O. performed theoretical and numerical analysis; S.B., M.K., C.S. and S.H. fabricated and characterized the semiconductor microcavity; E.A.O. and K.Y.B. wrote the paper with input from T.G., E.E. and T.C.H.L.; F.N., M.D.F., A.G.T, S.H., Y.Y. and Y.S.K. contributed to discussions and the shaping of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A.O. (elena.ostrovskaya@anu.edu.au).

METHODS

Experimental setup. The semiconductor sample used in the experiment is a GaAs/AlGaAs microcavity containing 12 quantum wells (QWs) (~13 nm wide each) sandwiched between distributed Bragg reflector mirrors (32/36 mirror pairs). To achieve the strong interaction regime between cavity photons and quantum-well excitons[1,31], the quantum wells are distributed in the sample via three sets of four located at the anti-nodes of the photon mode. The cavity photon mode is red-detuned by 2.8 meV from the exciton resonance at 1.546 eV, resulting in the exciton-polariton dispersion schematically shown in Fig. 1a. The sample is mounted on a cold finger inside a continuous flow microscopy cryostat and maintained at 5.6 K.

A schematic of the experimental apparatus is shown in Extended Data Fig. 1. The exciton-polariton condensate is formed by illuminating the sample by a quasi-continuous, off-resonant, linearly polarized pump beam derived from a continuous wave (CW) Ti:sapphire laser operating at 732 nm. The threshold power for the condensation is ~0.079 mW μm^{-2}. To minimize heating of the sample, the pump beam is chopped by an acoustic optical modulator (AOM). We use a digital micromirror device (DMD) to engineer the spatial pump profile in the shape of a Sinai billiard shown in Fig. 1b, which is then re-imaged onto the sample at normal incidence through a high numerical aperture (NA) microscope objective.

Owing to the continuous decay of the exciton-polaritons, coherent photons escape the cavity as a photoluminescence signal and carry all the information about the condensate[1]. The photoluminescence is then collected via the microscope objective and analysed using the CCD camera and spectrometer (Extended Data Fig. 1). We reconstruct the spatial modes by scanning the real space imaging across the slit of the spectrometer.

Creating exciton-polariton billiards. The DMD mirror is programmed to reflect the spatial pattern shown in Extended Data Fig. 2, thus creating a structured pump beam in the shape of a Sinai billiard[32]. The pump creates an inhomogeneous distribution of reservoir excitons in the plane of the quantum well, therefore inducing an effective potential for the condensed exciton-polaritons[10].

The two parameters of the billiard controlling the non-Hermitian dynamics of exciton-polaritons are the radius of the round corner (defect), R, and the thickness of the walls, d. The latter is different on the different sides of the perimeter due to the shape of the laser beam illuminating the DMD. Throughout the main text, we consider a continuous change of R ($0 < R/W < 1$), but only two modifications of d (shown in Extended Data Fig. 2).

We have verified that, for any R, when the thickness of the walls is varied within our experimental range, the pump power density remains approximately constant. For the 'thin' and 'thick' wall configurations shown in Extended Data Fig. 2, the values are 0.110 ± 0.0033 mW μm^{-2} (Extended Data Fig. 2a) and 0.117 ± 0.0047 mW μm^{-2} (Extended Data Fig. 2b), respectively. This effectively means that the height of the billiard potential walls, defined by the pump power, remains the same. Since the internal area and hence geometry of the billiard does not depend on d either, this leads us to conclude that the wall thickness controls mainly the imaginary part of the billiard potential.

Spectroscopy of the billiard. Above the condensation threshold, exciton-polaritons occupy multiple energy levels of the billiard potential, and in our experiments we comfortably resolve approximately the first 15 levels in the energy versus position spectrum. As the radius of the defect in the Sinai billiard grows, the area of the potential confining exciton-polaritons shrinks, so that the energy levels are blueshifted (see Fig. 1c, d). The spectral line profiles measured at fixed spatial positions in the vicinity of degeneracy highlighted in Fig. 1c, d are shown in Extended Data Fig. 3. The line profiles obtained for several values of the defect radius within the range ($0.4 < R/W < 0.65$) are plotted on the same plot and their relative blueshift is represented by the offset on the intensity axis.

Positions of the individual energy levels (Fig. 2a, b) for different values of R are derived from the spectroscopic peaks, as schematically shown in Extended Data Fig. 3, and the linewidths (Fig. 2c, d) are determined by the numerical fitting of the spectral profile. The errors indicated in Fig. 2 arise from the numerical fitting procedure and therefore are very small.

Modelling of the billiard. The full dynamics of the exciton-polariton condensate subject to off-resonant, incoherent optical pumping can be described by the generalized complex Gross–Pitaevskii (or Ginzburg–Landau) equation[33–35] for the condensate wavefunction, ψ:

$$i\hbar \frac{\partial \psi(\mathbf{r},t)}{\partial t} = \left[-\frac{\hbar^2}{2m}\nabla^2 + (g - i\gamma_{nl})|\psi|^2 + (g_R + i\hbar R)n_R(\mathbf{r}) - i\hbar\gamma \right]\psi + i\hbar\Re[\psi(\mathbf{r},t)]$$

(2)

Here m is the effective mass of the lower polariton, g is the polariton–polariton interaction strength, g_R is the strength of interaction between the reservoir and condensed polaritons, R is the rate of stimulated scattering into the condensed state, and γ is the spatially homogeneous decay rate of polaritons. The reservoir density distribution $n_R(\mathbf{r}) \propto P(\mathbf{r})$ is defined by the rate of reservoir (excitonic) polariton injection per unit area and time, $P(\mathbf{r})$. The parameter γ_{nl} entering equation (2) characterizes gain saturation and, in general, depends on the spatial distribution of the pump. In our numerical calculations, we take the γ_{nl} to be small and spatially homogeneous due to the weak overlap between the condensate and the pumping area.

The model, equation (2), was initially suggested phenomenologically[33] and subsequently derived from the semiclassical Maxwell–Bloch equations[34]. It qualitatively coincides with the generalized open-dissipative Gross–Pitaevskii model[36] augmented with the rate equation for the excitonic reservoir density:

$$\frac{\partial N_R}{\partial t} = P(\mathbf{r}) - (\gamma_R + R|\psi|^2)N_R$$

in the regime of near-threshold pumping[37]. In this limit, the steady state reservoir density distribution can be expressed as $N_R \approx P(\mathbf{r})/\gamma_R - RP(\mathbf{r})|\psi|^2/\gamma_R^2 = n_R(\mathbf{r}) - \gamma_{nl}\hbar^{-1}R^{-1}|\psi|^2$, where γ_R is the decay rate of reservoir polaritons.

The phenomenological energy relaxation[35,38,39], which is essential to adequately model the multi-mode nature of the condensate[38], is taken in the following form[35,38]:

$$\Re[\psi(\mathbf{r},t)] = \alpha n_R \left[\mu(\mathbf{r},t) - i\hbar\frac{\partial}{\partial t} \right]\psi(\mathbf{r},t)$$

where α is the energy relaxation rate, and $\mu(\mathbf{r}, t)$ is a local chemical potential of the condensate.

We use equation (2) to obtain the structure of the spatial modes of the exciton-polariton condensate corresponding to peaks of the energy spectrum. The parameters of the model used for our dynamical simulations are as follows: $m = 5 \times 10^{-5}\,m_e$, where m_e is the free electron mass, $g = 2 \times 10^{-3}$ meV μm^2, $g_R = 2g$, $\hbar R = 6 \times 10^{-4}$ meV μm^2, $\gamma = 0.1$ ps^{-1}, $\gamma_{nl} = 0.3g$, $\alpha = 1.2 \times 10^{-3}$ μm^2 ps^{-1} meV^{-1}. The effective potential height is max(V') = 2.25 meV, and the billiard wall profile given by the reservoir density distribution, $n_R(\mathbf{r})$, is convoluted with a Gaussian profile to account for the 'soft' edges of the potential created by the optical excitation and exciton diffusion.

The spatial modes computed numerically using the fully nonlinear, open-dissipative dynamical model, equation (2), are presented in the bottom row of Extended Data Fig. 4. For comparison, the middle row of Extended Data Fig. 4 shows the single-particle eigenstates of the complex linear potential induced by the excitonic reservoir: $V(\mathbf{r}) = V' + iV'' \equiv g_R n_R(\mathbf{r}) + i\hbar[Rn_R(\mathbf{r}) - \gamma]$, with both real and imaginary parts V', V'' proportional to the pumping rate[36], $P(\mathbf{r})$. One can see that the condensate dynamics described by equation (2) effectively populates the eigenstates of the linear complex effective potential. The validity of our model is confirmed by the excellent agreement with the experimental images of the billiard modes presented in the top row of Extended Data Fig. 4.

In agreement with previous studies[40], the nonlinearity due to exciton-polariton interactions strongly determines the relative population of the eigenstates, as well as the overall blueshift of the eigenenergies. The eigenenergies are complex, and so the spectral linewidths may exceed the level separation. For this reason, in our experiment some of the higher-order energy-filtered wavefunctions represent superpositions of neighbouring eigenstates. For example, the seventh mode measured in the experiment (last column in Extended Data Fig. 4) is, in fact, a superposition of eigenstates eight and nine, as revealed by the comparison with the numerically calculated modes. In contrast, the lower-order modes in Extended Data Fig. 4 represent almost pure eigenstates, having a very weak (less that 10%) admixture of the neighbouring eigenstates.

Hybridization of modes. Hybridization of modes occurs in the vicinity of crossing and anti-crossing of the energy levels in Fig. 2a and b. In these regions, the billiard modes are different in shape to the uncoupled modes away from the (near-)degeneracy. In experiments, it is hard to spectrally resolve pure modes in the hybridization region since their spectral linewidths exceed the peak separation. Therefore, what is experimentally imaged and shown in the insets of Fig. 2a and b is a superposition of two modes. This is especially true for Fig. 2b, where the spectral peaks (but not the linewidths) precisely coincide at the crossing, so that in the experiment we can only image a single mode corresponding to a single peak.

To match the spatial distributions obtained in the experiment with those found numerically, we plot superpositions of the pure eigenstates found in numerical simulations: $\varphi_s = \alpha\varphi_3 + \beta\varphi_4 e^{i\vartheta}$, where $\varphi_{3,4}$ are the pure eigenstates 3 and 4, α and β are their relative amplitudes, and ϑ is the relative phase. We find that only the relative phase $\vartheta = \pi/2$ can produce a superposition that fits well with the experiment. These spatial modes (pure and superposition states) for the thick and thin billiard in the hybridization region (anti-crossing and crossing of eigenenergies, respectively) are shown in Extended Data Fig. 5.

Note that this mode mixing is performed only in the hybridization region. Away from this region, the experimentally imaged and numerically calculated modes

match extremely well. Importantly, in Fig. 4, in order to perform a reliable phase extraction from numerically found modes away from the (near-)degeneracy, it is absolutely necessary to trace the continuous variation of phase of the pure modes as we pass the hybridization region. For this reason, we did not mix numerically found pure modes to match experimentally imaged spatial distributions of super-position states. This explains visible discrepancies between the spatial structure of numerically calculated and experimentally imaged modes in the hybridization regions in Fig. 4.

Coupled-mode model. The behaviour of any two non-Hermitian modes of the billiard potential near the degeneracy point can be described by a standard coupled-mode model written in the dimensionless form as follows:

$$i\frac{\partial \psi_{n,n'}(\boldsymbol{r},t)}{\partial t} = \left[-\nabla^2 + V'(\boldsymbol{r}) + iV''(\boldsymbol{r})\right]\psi_{n,n'} + \Omega\psi_{n',n} \qquad (3)$$

where Ω characterizes the coupling strength between the states n and n'. Separating the temporal and spatial dependence of the wavefunctions, $\psi_n = a_n(t)\varphi_n(\boldsymbol{r})$, substituting this ansatz into equation (3), and integrating out the spatial degrees of freedom, leads us to the eigenvalue equation (1) in the main text, where $(n,n') = (1,2)$. The real energies of the modes away from the degeneracy in equation (1) are defined by the shape of the billiard potential, $\left[-\nabla^2 + V'(\boldsymbol{r})\right]\varphi_n(\boldsymbol{r}) = E_n\varphi_n(\boldsymbol{r})$, the complex parts of the eigenenergies are given by the overlap between the billiard modes and the exciton reservoir, $\Gamma_n \propto \int V''(\boldsymbol{r})|\varphi_n(\boldsymbol{r})|^2 \mathrm{d}^2\boldsymbol{r}$, and the off-diagonal matrix elements in equation (1) are determined by the degree of spatial overlap between the two modes, $q \propto \int \varphi_n^*(\boldsymbol{r})\varphi_{n'}(\boldsymbol{r})\,\mathrm{d}^2\boldsymbol{r}$. Here we assume that the uncoupled modes are properly normalized.

We stress that our theory essentially uncovers the non-Hermitian features of degeneracies of the modes of the entire billiard, which can be considered as a two-dimensional resonator in the plane of the quantum well. This approach is conceptually close to, for example, the theory of two-dimensional PT-symmetric systems[41] and differs from the well developed coupled-mode theories for weakly coupled non-Hermitian single-mode resonators.

31. Kavokin, A., Baumberg, J., Malpuech, G. & Laussy, F. *Microcavities* (Oxford Univ. Press, 2007).
32. Sinai, Y. G. Dynamical systems with elastic reflections. *Russ. Math. Surv.* **25,** 137–189 (1970).
33. Keeling, J. & Berloff, N. G. Spontaneous rotating vortex lattices in a pumped decaying condensate. *Phys. Rev. Lett.* **100,** 250401 (2008).
34. Cristofolini, P. *et al.* Optical superfluid phase transitions and trapping of polariton condensates. *Phys. Rev. Lett.* **110,** 186403 (2013).
35. Askitopoulos, A. *et al.* A robust platform for engineering pure-quantum-state transitions in polariton condensates. *Phys. Rev. B* **92,** 035305 (2015).
36. Wouters, M. & Carusotto, I. Excitations in a nonequilibrium Bose–Einstein condensate of exciton polaritons. *Phys. Rev. Lett.* **99,** 140402 (2007).
37. Smirnov, L. A., Smirnova, D. A., Ostrovskaya, E. A. & Kivshar, Yu. S. Dynamics and stability of dark solitons in exciton-polariton condensates. *Phys. Rev. B* **89,** 235310 (2014).
38. Wouters, M. Energy relaxation in the mean-field description of polariton condensates. *New J. Phys.* **14,** 075020 (2012).
39. Wertz, E. *et al.* Propagation and amplification dynamics of 1D polariton condensates. *Phys. Rev. Lett.* **109,** 216404 (2012).
40. Eastham, P. R. Mode locking and mode competition in a nonequilibrium solid-state condensate. *Phys. Rev. B* **78,** 035319 (2008).
41. Ge, L. & Stone, A. D. Parity-time symmetry breaking beyond one dimension: the role of degeneracy. *Phys. Rev. X* **4,** 031011 (2014).
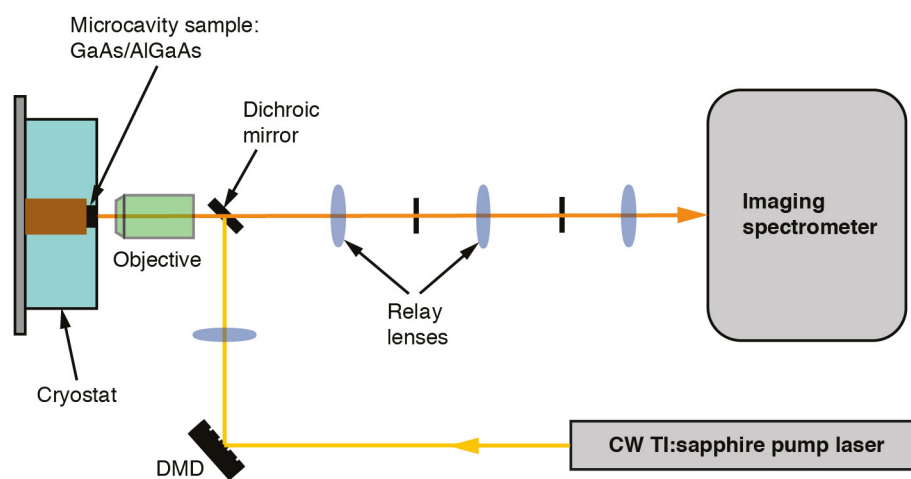
**Extended Data Figure 1 | Diagram of the experimental apparatus.** See Methods for details.

a

3.4 μm

14 μm

23 μm

4.5 μm

b

5.2 μm

14 μm

7.2 μm
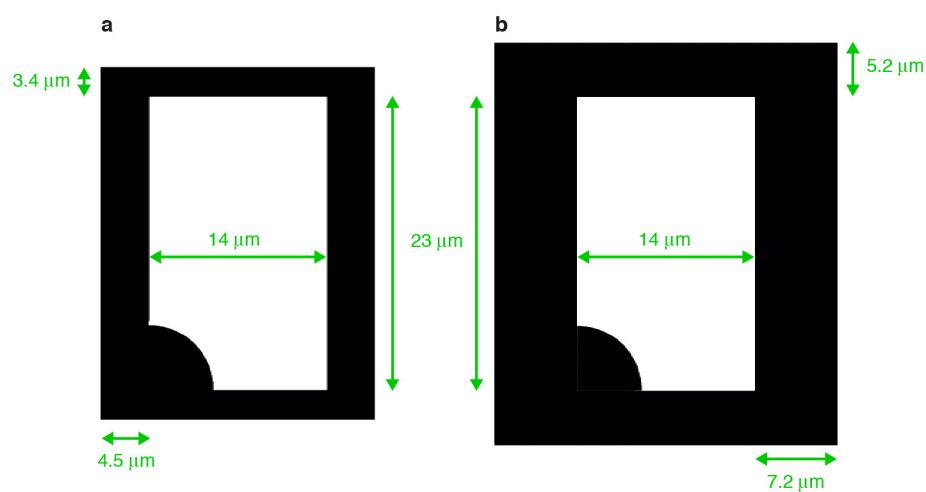
Extended Data Figure 2 | Schematics of the optically induced billiard potential with two different wall thicknesses. a, Thin walls; **b**, thick walls. The active regions corresponding to the optical pump are shown in black, and we note that the enclosed area does not change with wall thickness.
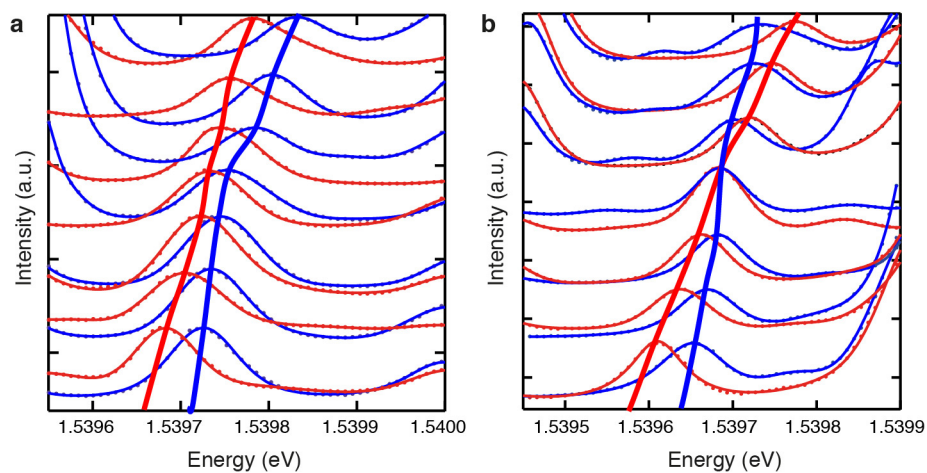
Extended Data Figure 3 | Effect of wall thickness on spectroscopic line profiles of the Sinai billiard. a, b, Profiles are shown in the vicinity of the degeneracy for the levels highlighted in Fig. 1c, d with thick (**a**) and thin (**b**) walls. The thick lines demonstrate the principle of data extraction for anti-crossing (**a**) and crossing (**b**) of the energy levels corresponding to those shown in Fig. 2a and b, respectively
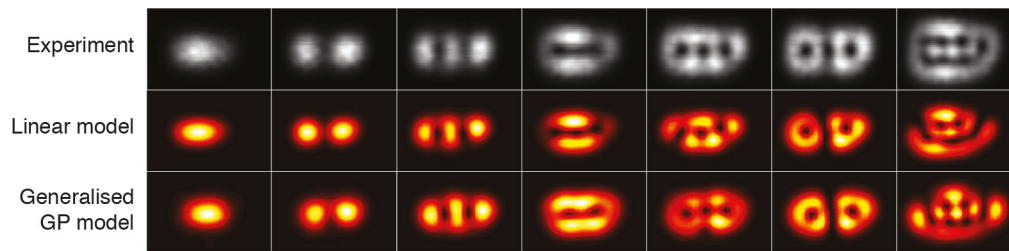
Extended Data Figure 4 | Spatial density distribution of the first seven simultaneously populated lowest-energy modes of the Sinai billiard. Spatial density distributions were obtained from the thick-wall setup (Extended Data Fig. 2b) with $R/W = 0.35$. Top row, experimentally imaged; middle row, calculated using the effective linear potential model; bottom row, calculated using the full dynamical model given by equation (2).
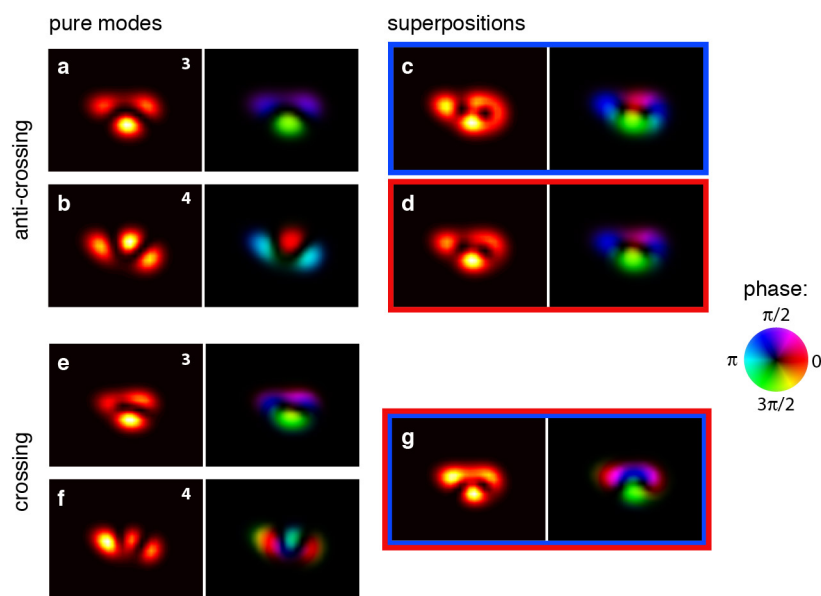
Extended Data Figure 5 | Spatial modes in the hybridization regions.
a–g, Calculated spatial modes; each panel shows the modulus squared of the wavefunction (left) and the wavefunction's phase distribution (right, colour coded). **a**, **b**, **e**, **f**, Numerically calculated pure spatial eigenstates (modes 3 (**a**, **e**) and 4 (**b**, **f**)) for the Sinai billiard with thick and thin walls in the corresponding hybridization regions shown in Fig. 2a and b, respectively.

c, **d**, **g**, The superpositions of modes 3 and 4 that match the experimentally imaged modes shown in Fig. 2; **c** (boxed in blue) and **d** (boxed in red) correspond to the blue and red curves of Fig. 2a, respectively, while **g** (boxed in red and blue) corresponds to the crossing point in Fig. 2b. The relative populations of pure modes in the superposition states are: **c**, $|\alpha|^2 = 0.85$ and $|\beta|^2 = 0.15$; **d**, $|\alpha|^2 = 0.65$ and $|\beta|^2 = 0.35$; **g**, $|\alpha|^2 = 0.60$ and $|\beta|^2 = 0.40$.

LETTER

The vulnerability of Indo–Pacific mangrove forests to sea–level rise

Catherine E. Lovelock[1,2], Donald R. Cahoon[3], Daniel A. Friess[4], Glenn R. Guntenspergen[3], Ken W. Krauss[5], Ruth Reef[1,2,6], Kerrylee Rogers[7], Megan L. Saunders[2], Frida Sidik[8], Andrew Swales[1,9], Neil Saintilan[10], Le Xuan Thuyen[11] & Tran Triet[11,12]

Sea-level rise can threaten the long-term sustainability of coastal communities and valuable ecosystems such as coral reefs, salt marshes and mangroves[1,2]. Mangrove forests have the capacity to keep pace with sea-level rise and to avoid inundation through vertical accretion of sediments, which allows them to maintain wetland soil elevations suitable for plant growth[3]. The Indo-Pacific region holds most of the world's mangrove forests[4], but sediment delivery in this region is declining, owing to anthropogenic activities such as damming of rivers[5]. This decline is of particular concern because the Indo-Pacific region is expected to have variable, but high, rates of future sea-level rise[6,7]. Here we analyse recent trends in mangrove surface elevation changes across the Indo-Pacific region using data from a network of surface elevation table instruments[8–10]. We find that sediment availability can enable mangrove forests to maintain rates of soil-surface elevation gain that match or exceed that of sea-level rise, but for 69 per cent of our study sites the current rate of sea-level rise exceeded the soil surface elevation gain. We also present a model based on our field data, which suggests that mangrove forests at sites with low tidal range and low sediment supply could be submerged as early as 2070.

Intertidal mangrove forests occur on tropical and subtropical shorelines, and provide a wide range of ecosystem services, including the support of fisheries, coastal protection and carbon sequestration, which are collectively and conservatively estimated to be worth US$194,000 per hectare per year (refs 11, 12). Although mangrove tree species are able to tolerate inundation by tides, they can die and their former habitat can convert to open water or tidal flats when sea-level rise (SLR) causes the frequency and duration of inundation to exceed species-specific physiological thresholds[13], resulting in shoreline retreat[14]. In low-sediment-supply systems such as Caribbean atolls, the capacity of the soil surface to keep pace with SLR is strongly dependent on the accumulation of organic matter derived from roots that decompose slowly in anaerobic soils[15]. But sediment accretion on the soil surface in the Indo-Pacific region can also play a crucial role in surface elevation gains[16].

Changes in the elevation of the soil surface over time can be measured using the surface elevation table–marker horizon (SET–MH) methodology[8,9], which has been widely used and recommended for monitoring intertidal surface-elevation trajectories in coastal wetlands[10]. Here we use an extensive network of SET–MH stations (Fig. 1) with records of 1–16.6 years in length to investigate the role of sediments in maintaining surface elevation gain in these Indo-Pacific mangrove forests and to identify their vulnerability to future SLR. Recent trends in mangrove surface elevation change across 27 sites in the Indo-Pacific (Supplementary Table 1) were analysed with respect to environmental factors, including suspended-matter concentration and the regional rate of SLR obtained from tide gauges. Future

vulnerability to SLR was modelled on the basis of the results of this analysis using a surface elevation change model and likely future SLR scenarios.

Throughout the Indo-Pacific region, we found that mangrove soil-surface elevation gains are strongly dependent on rates of accretion of sediment on the soil surface (Fig. 2a) as well as subsurface organic matter accumulation, which has been observed in sites in the Caribbean[15]. One site in southeast Java, Indonesia, has particularly high rates of surface accretion, owing to a mud-volcano eruption[17], but even with this site removed from the analysis, surface elevation gain remains significantly correlated with sediment accretion ($R^2 = 0.259$, $P < 0.001$, F test). As expected from theoretical models[18], we found that the concentration of total suspended matter (TSM) in the water column, derived from remotely sensed MERIS (medium resolution imaging spectrometer) imagery, was proportional to surface accretion (Fig. 2b) and to surface elevation gains (Fig. 2c), although the relationship between surface elevation and TSM was more variable than that observed between surface elevation and locally measured rates of surface accretion. These relationships link the supply of sediments to the maintenance of soil elevation relative to sea level in mangrove forests at regional scales within the Indo-Pacific region. Other factors (such as rate of SLR, geomorphology, habitat and dominant species) explained a smaller proportion of the variation in the surface elevation gains (Extended Data Table 1). On the basis of our network of SET–MH sites, we conclude that sediment supply is important to surface elevation gains and therefore to preventing mangrove-forest loss in the future.

We found that 69% of surface elevation records in the Indo-Pacific data set (90 out of a total of 153 SET–MH stations) had rates of surface elevation gain that were less than the long-term rate of SLR for the region (Extended Data Fig. 1b). The remaining 31% of the records are from sites in Australia, New Zealand, Vietnam and Indonesia. Many of the sites that had rates of surface elevation gain less than SLR also exhibited shallow subsidence (Extended Data Fig. 1a). Shallow subsidence can be caused by a range of factors that increase compaction of the near-surface sediments and that are responsive to local environmental factors, including forest degradation[19]. But whether subsidence and the 'elevation deficit' relative to local rates of SLR indicate vulnerability of these mangrove forests to loss with increasing rates of SLR is unknown. If the topography allows the mangrove forest to migrate landward, with no anthropogenic barriers (such as infrastructure or flood-defence barriers), then mangroves may delay submergence by 'back-stepping' into adjacent habitats[20]. However, barriers to landward expansion of mangrove forests occur throughout the Indo-Pacific region, particularly in sites that have intensive aquaculture, urban development and low-lying agricultural land. We have therefore assumed that broad-scale landward retreat of human settlements in

[1]School of Biological Sciences, The University of Queensland, Brisbane 4072, Australia. [2]Global Change Institute, The University of Queensland, Brisbane 4072, Australia. [3]Patuxent Wildlife Research Center, United States Geological Survey, Maryland 20708, USA. [4]Department of Geography, National University of Singapore, 1 Arts Link, Singapore 117570, Singapore. [5]National Wetlands Research Center, United States Geological Survey, Louisiana 70506, USA. [6]Cambridge Coastal Research Unit, Department of Geography, University of Cambridge, Downing Place, Cambridge CB2 3EN, UK. [7]School of Earth and Environmental Science, University of Wollongong, Wollongong 2522, Australia. [8]The Institute for Marine Research and Observation, Ministry of Marine Affairs and Fisheries, Bali 82251, Indonesia. [9]National Institute of Water and Atmospheric Research, Hamilton 3251, New Zealand. [10]Department of Environmental Sciences, Macquarie University, Sydney 2109, Australia. [11]University of Science, Vietnam National University, Ho Chi Minh City, Vietnam. [12]International Crane Foundation, Wisconsin 53913, USA.
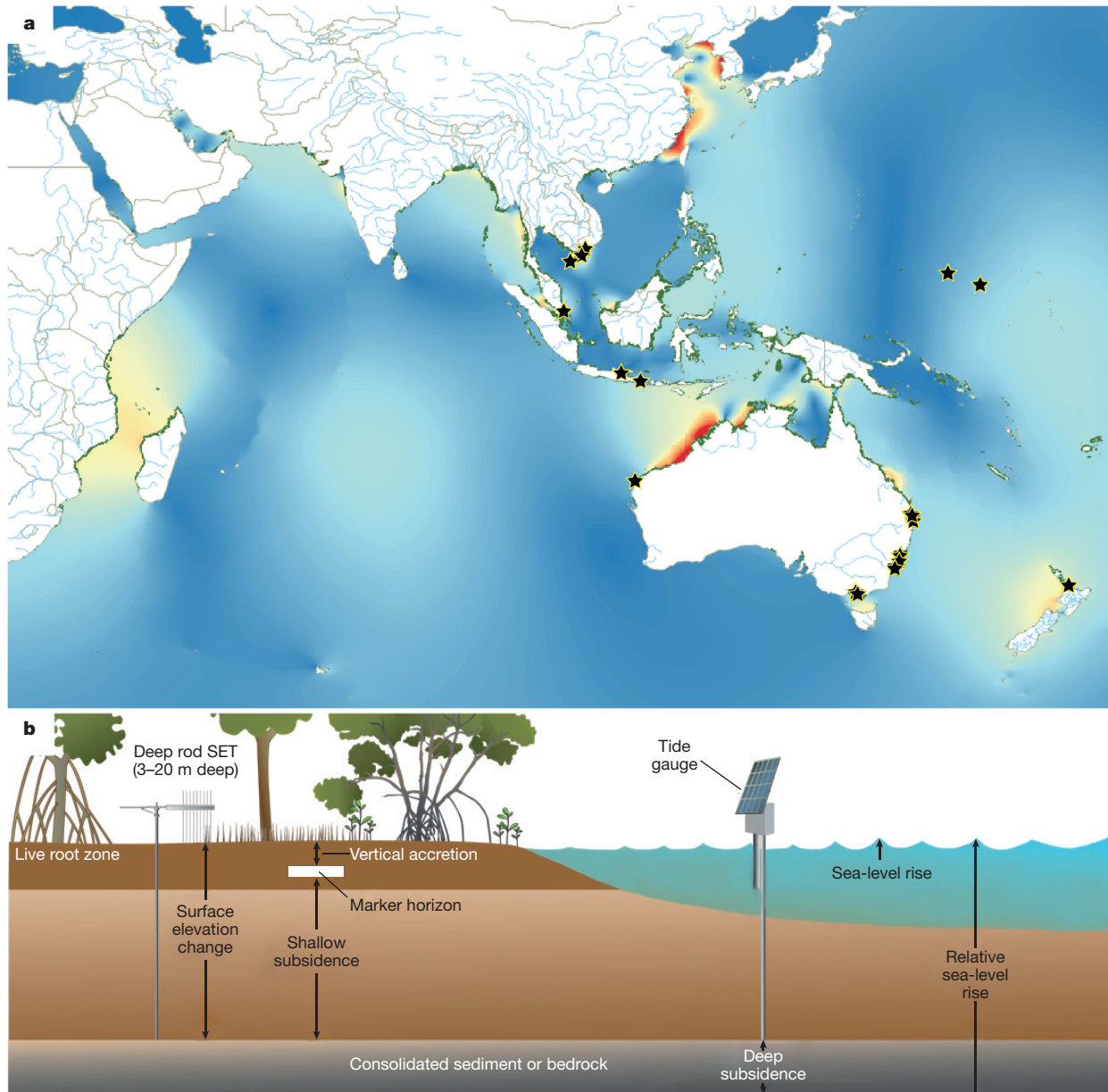
Figure 1 | Map of the Indo-Pacific region study sites and a schematic of the SET–MH. a, Study sites are indicated by stars; mangrove forests are shown in dark green. The colour of the coastal ocean represents variation in tidal range (Aviso + FES2012 tide model), where blue is microtidal (0–2 m), yellow is mesotidal (2–4 m) and red is macrotidal (>4 m). **b**, The SET–MH installation monitors changes in soil-surface elevation, surface accretion above a marker horizon and shallow subsidence (by difference[8,9]); see Methods for details.

the region is unlikely as a result of political uncertainty and because in many nations coastal inhabitants are 'trapped' by a lack of capital and available inland sites that would support migration[21].

To examine the future vulnerability of Indo-Pacific mangroves to SLR, we developed a model of mangrove habitat suitability based on position in the tidal frame. Mangrove forests persist in the portion of the tidal frame from mean sea level (MSL) to the level of the highest astronomical tide, which generally corresponds to the highest elevation at which mangroves can survive[22]. This gives rise to what is termed a wetland's 'elevation capital', or the potential of an intertidal wetland to remain within a suitable inundation regime at that site (that is, above MSL) despite subsiding relative to local SLR[23]. For example, mangrove forests occupying high intertidal sites that have a 10-m tidal range (such as the Kimberly coast of Australia) would need to lose up to 5 m of elevation capital to reach MSL. In contrast, high intertidal sites with a tidal range of 1 m (such as the Caribbean and parts of Indonesia)

would have to lose only 0.5 m of elevation to put the entire contemporary forest at or below MSL. Assuming that mangrove forest species cannot persist below approximately MSL, we estimate the time to inundation and thus loss of the forest by using tidal range as a surrogate for the elevation capital within the ecosystem.

Over the range of elevation deficits within our data set, we estimated the time until complete submergence of the forest at sites with varying tidal range (and thus varying elevation capital). This model, which subtracts elevation from the elevation capital over time, assumes constant rates of SLR. Assuming an elevation deficit of 20 mm yr^{-1} (that is, sea level rising 20 mm yr^{-1} faster than mangrove surface elevation gain), which occurs at some of our sites owing to high local rates of SLR and shallow subsidence (for example, Indonesia), we project complete submergence of the forests in 100 years wherever tidal ranges are less than 4 m (Extended Data Fig. 2). At an elevation deficit of 6 mm yr^{-1} (the mean elevation deficit for our sites with elevation deficits), we
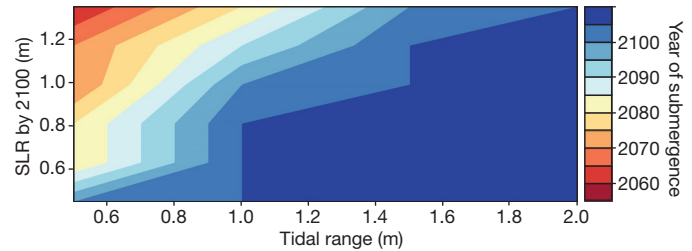
a

b

c

Figure 3 | Year in which mangroves are predicted to be submerged at sites with low (<2.5 g m^{-3}) sediment availability over variation in tidal range and rates of SLR. The darkest blue region indicates no submergence predicted within the modelling time frame (until 2100). At high sediment supply (>2.5 g m^{-3}), mangrove forests were not predicted to be submerged by 2100. See Methods for further details.

Figure 2 | The relationship between mangrove soil-surface elevation gains and sediment availability. a–c, Relationships between soil-surface elevation gains and accretion on the soil surface (**a**), accretion on the soil surface and average annual TSM derived from MERIS satellite imagery (**b**), and surface elevation gains and average annual TSM (**c**). Data points are coloured as follows: pink, Indonesia; dark green, Vietnam; light green, New Zealand; yellow, western Australia; dark blue, Micronesia; grey, Singapore; white, eastern Australia. Solid lines are linear regressions: **a**, (surface elevation gain) = $(-4.44 \pm 0.95) + (0.78 \pm 0.03) \times$ (surface accretion), $R^2 = 0.849$, $P < 0.0001$, F test (for overall significance of the linear regression); **b**, (surface accretion) = $(4.15 \pm 1.08) + (1.57 \pm 0.15) \times$ TSM, $R^2 = 0.443$, $P < 0.0001$, F test (excluding data from Porong, Indonesia); **c**, (surface elevation gain) = $(1.38 \pm 0.83) + (0.51 \pm 0.11) \times$ TSM, $R^2 = 0.122$, $P < 0.0001$, F test (excluding data from Porong, Indonesia); the indicated uncertainties are standard errors. Source Data for this figure are available online.

estimate it would take 100–300 years for high intertidal forests to be lower than MSL, while at low elevation deficits (1 mm yr^{-1}) the forests may persist for thousands of years. The palaeorecord is consistent with high levels of persistence of mangroves through time when rates of SLR are low to moderate (that is, low levels of elevation deficits). For example, in Belize there is evidence that mangrove forests persisted

for long intervals over the Holocene epoch during periods when rates of SLR were less than 5 mm yr^{-1} (ref. 15). Additionally, there is evidence that high intertidal mangrove forests in northern Australia persisted for thousands of years despite relatively high rates of SLR[14]. However, evidence of overwhelming flooding and loss of mangrove forests is also evident during past rapid rises in sea level[14].

To synthesize the effects of sediment supply and accelerating rates of SLR (and corresponding elevation deficits) on the fate of mangrove forests, we formulated a second model that assessed the probable time to submergence of mangrove forests over the range of observed rates of surface elevation gains with no landward migration and over a range of tidal amplitudes. According to the model, mangrove forests are likely to persist at sites with high tidal range even with high rates of SLR and low levels of sediment availability (Fig. 3), consistent with palaeo-observations[14] and theory[16,18]. However, at sites with low tidal range, forests will be vulnerable by 2080 at moderate SLR (0.8 m by 2100).

We cannot estimate the absolute extent of losses of mangrove cover over the region because measurements of mangrove forest elevation in the region are too coarse; however, our model provides a semi-empirical indication of the conditions under which mangrove loss is likely with SLR and locations where management of sediment supply and space for landward migration are vital to ensure mangrove forests survive into the future. Our model indicates that the outlook for mangrove forests in some locations is poor under relatively low rates of SLR—the Intergovernmental Panel on Climate Change (IPCC) Representative Concentration Pathway 6 (RCP6) scenario—with submergence of mangroves by 2070 predicted in the Gulf of Thailand, the southeast coast of Sumatra, the north coasts of Java and Papua New Guinea and the Solomon Islands (Fig. 4). In contrast, the outlook for the persistence of mangroves into the future is more positive in east Africa, the Bay of Bengal, eastern Borneo and northwestern Australia, where there are relatively large tidal ranges and/or higher sediment supply.

Our model does not account for long-term and nonlinear feedbacks within the system where elevation deficits may be enhanced or reduced, for example, through episodic high-wave-energy events that cause erosion[24], degradation of forests[25], other stochastic events such as intense storms that may alter hydrology or deliver sediment pulses[26], or changes in ocean circulation that may influence regional rates of SLR[7]. The frequency and intensity of these events are predicted to increase under climate-change scenarios[2], and all of these factors will influence the length of time before forest submergence and loss. Our model also does not include subsidence (or uplift) that occurs below the SET benchmark[9], which in some locations may strongly influence the time until submergence. But shifts in the way sediment is managed, and reversing forest degradation and thus enhancing organic matter inputs to sediments may extend the persistence of mangroves for hundreds of years (for example, reducing elevation deficits by 6 mm yr^{-1} extended the time until submergence from 83 years to 167 years for sites with a 2-m tidal range). In coastal and estuarine systems with reduced upstream sediment inputs due to human modifications[27], the potential

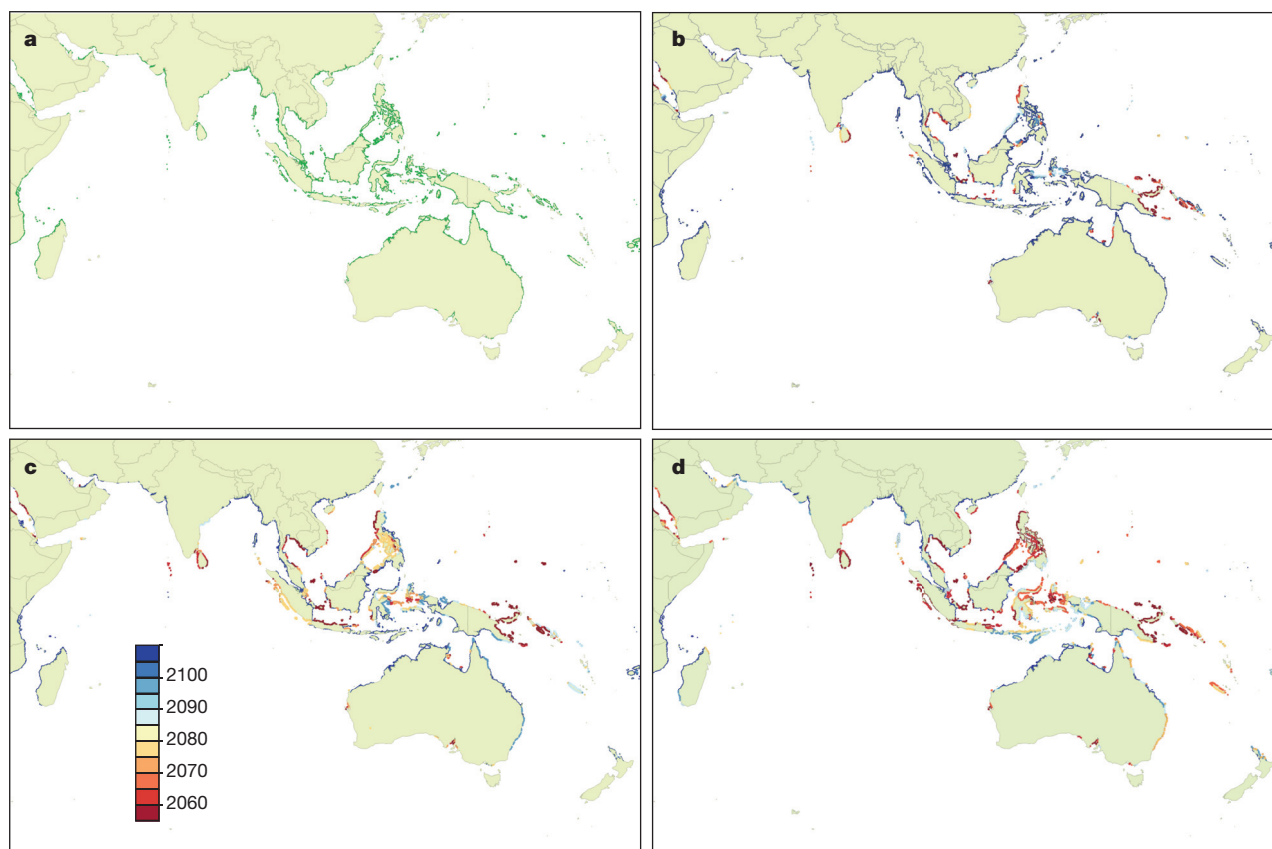**Figure 4 | Mangrove forest distribution in the Indo-Pacific region. a**, Dark green areas indicate current mangrove forests. **b–d**, Predicted decade of mangrove forest submergence, indicated by the colour scale in **c**, for IPCC RCP6 (0.48 m SLR by 2100) (**b**), RCP8.5 (0.63 m SLR by 2100) (**c**), and a more extreme scenario (1.4 m SLR by 2100) (**d**). The darkest blue region indicates no submergence predicted within the modelling time frame (until 2100). The model assumes landward migration of mangrove forests is not possible.

for eco-geomorphic feedbacks that delay the onset of mangrove-forest loss is diminished.

Data from our network of sites indicate that the fate of mangroves in the Indo-Pacific with SLR is strongly linked to the availability of suspended matter, which is important for increases in soil-surface elevation and enables mangroves to maintain their elevation within the tidal frame above MSL. The importance of sediment supply for the resilience of mangrove forests in the face of SLR has been inferred from the palaeontological record[14] and from recent observed changes in mangrove coasts where sediment supply has been reduced, owing to damming of rivers[27]. In Thailand, there has been an 80% reduction of sediment supply in the Chao Phraya River delta, which, in combination with surface subsidence caused by groundwater extraction, has resulted in kilometres of mangrove shoreline retreat[28]. Within the Mekong River system, planned construction of dams and reductions in sediment supply[28] will have a devastating effect on the local coastal sediment budget and the long-term persistence of mangrove forests. Management of the coast and particularly of the river systems that deliver much of the sediment to the region is therefore vital for the survival of mangrove forests. Although sediment supply at some sites may be maintained as a legacy of prior forest clearing of catchments (which leads to erosion of soil), the restriction of sediment supply caused by the building of dams is a major issue that will contribute to mangrove losses in the future.

More than half the mangrove forests we studied have already lost elevation relative to sea level. With constant rates of SLR where tidal ranges are large and sediment supplies are maintained, mangrove forests in the upper intertidal zone may survive thousands of years before they are threatened by submergence. However, under moderate emissions scenarios (for example, IPCC RCP6) at sites with low tidal ranges and low sediment supply, mangrove forests may be lost by 2080. Our work emphasizes the urgent need to plan for the maintenance of sediment supply in river systems that are expected to be heavily modified and dammed in the future, to reverse forest degradation that reduces organic matter inputs and to plan for the landward migration of mangrove forests to higher elevations in locations where sediment supply is expected to be restricted.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Nicholls, R. J. & Cazenave, A. Sea-level rise and its impact on coastal zones. *Science* **328,** 1517–1520 (2010).
2. Woodruff, J. D., Irish, J. L. & Camargo, S. J. Coastal flooding by tropical cyclones and sea-level rise. *Nature* **504,** 44–52 (2013).
3. Kirwan, M. L. & Megonigal, J. P. Tidal wetland stability in the face of human impacts and sea-level rise. *Nature* **504,** 53–60 (2013).
4. Giri, C. *et al.* Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecol. Biogeogr.* **20,** 154–159 (2011).
5. Milliman, J. D. & Farnsworth, K. L. *River Discharge to the Coastal Ocean: A Global Synthesis* (Cambridge Univ. Press, 2011).
6. Church, J. A. *et al.* in *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Ch. 13 (Cambridge Univ. Press, 2013).
7. Stammer, D. *et al.* Causes for contemporary regional sea level changes. *Annu. Rev. Mar. Sci.* **5,** 21–46 (2013).
8. Cahoon, D. R. *et al.* High-precision measurements of wetland sediment elevation: I. Recent improvements to the sedimentation-erosion table. *J. Sediment. Res.* **72,** 730–733 (2002).

9. Callaway, J. C., Cahoon, D. R. & Lynch, J. C. in *Methods in Biogeochemistry of Wetlands* (eds DeLaune, R. D. *et al.*) 901–917 (Soil Science Society of America, 2013).
10. Webb, E. L. *et al.* A global standard for monitoring coastal wetland vulnerability to accelerated sea-level rise. *Nature Clim. Change* **3,** 458–465 (2013).
11. Costanza, R. *et al.* Changes in the global value of ecosystem services. *Global Environ. Change* **26,** 152–158 (2014).
12. Brander, L. M. *et al.* Ecosystem service values for mangroves in Southeast Asia: a meta-analysis and value transfer application. *Ecosyst. Services* **1,** 62–69 (2012).
13. Ball, M. C. Ecophysiology of mangroves. *Trees Struct. Funct.* **2,** 129–142 (1988).
14. Woodroffe, C. D. Response of tide-dominated mangrove shorelines in Northern Australia to anticipated sea-level rise. *Earth Surf. Proc.. Land.* **20,** 65–85 (1995).
15. McKee, K. L., Cahoon, D. R. & Feller, I. C. Caribbean mangroves adjust to rising sea level through biotic controls on change in soil elevation. *Global Ecol. Biogeogr.* **16,** 545–556 (2007).
16. Kirwan, M. L. & Murray, A. B. A coupled geomorphic and ecological model of tidal marsh evolution. *Proc. Natl Acad. Sci. USA* **104,** 6118–6122 (2007).
17. Jennerjahn, T. C. *et al.* Environmental impact of mud volcano inputs on the anthropogenically altered Porong River and Madura Strait coastal waters, Java, Indonesia. *Estuar. Coast. Shelf Sci.* **130,** 152–160 (2013).
18. Fagherazzi, S. *et al.* Numerical models of salt marsh evolution: ecological, geomorphic, and climatic factors. *Rev. Geophys.* **50,** RG1002 (2012).
19. Krauss, K. W. *et al.* How mangrove forests adjust to rising sea level. *New Phytol.* **202,** 19–34 (2014).
20. Saintilan, N. *et al.* Mangrove expansion and salt marsh decline at mangrove poleward limits. *Global Change Biol.* **20,** 147–157 (2014).
21. Black, R., Bennett, S. R. G., Thomas, S. M. & Beddington, J. R. Climate change: migration as adaptation. *Nature* **478,** 447–449 (2011).
22. Alongi, D. M. Present state and future of the world's mangrove forests. *Environ. Conserv.* **29,** 331–349 (2002).
23. Cahoon, D. R. & Guntenspergen, G. R. Climate change, sea-level rise, and coastal wetlands. *Nat. Wetl. Newslett.* **32,** 8–12 (2010).
24. Winterwerp, J. C. *et al.* Defining eco-morphodynamic requirements for rehabilitating eroding mangrove-mud coasts. *Wetlands* **33,** 515–526 (2013).
25. Lang'at, J. K. S. *et al.* Rapid losses of surface elevation following tree girdling and cutting in tropical mangroves. *PLoS ONE* **9,** e107868 (2014).
26. Cahoon, D. R. A review of major storm impacts on coastal wetland elevations. *Estuar. Coast.* **29,** 889–898 (2006).
27. Giosan, L., Syvitski, J., Constantinescu, S. & Day, J. Climate change: protect the world's deltas. *Nature* **516,** 31–33 (2014).
28. Kondolf, G. M., Rubin, Z. K. & Minear, J. T. Dams on the Mekong: cumulative sediment starvation. *Water Resour. Res.* **50,** 5158–5169 (2014).

Author Contributions All authors participated in a collaborative workshop or contributed field data, contributed to the conceptualization of models and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.E.L. (c.lovelock@uq.edu.au).

METHODS

SET–MH method description. The SET and the later-developed rod-SET consist of a benchmark rod driven in sections through the soil profile to resistance, often to 10–25-m depth in the soil or to when bedrock is reached (Fig. 1b). After installation of the benchmark rod, a portable horizontal arm is attached, and fixed points (usually four positions around the top of the rod) are used to measure the distance to the substrate surface using a series of vertical pins lowered to the soil surface (Fig. 1b). Total surface height measurements have confidence intervals of ±1.3 mm (ref. 8). SET data are usually complemented with monitoring of accretion on the soil surface using artificial soil marker horizons typically made of feldspar, sand or other resistant material, which simultaneously allows users to quantify rates of vertical surface accretion (that is, sediment deposition; Fig. 1b). The complete SET–MH installation provides observations of net surface elevation change above the benchmark depth as well as accretion on the surface of the wetland. These values may be compared to infer whether surface or subsurface processes are contributing to surface elevation gains. For example, if accretion on the soil surface is equivalent to surface elevation gain, then accretion on the soil surface, whether of mineral or organic origin, is the major process contributing to elevation gain. However, if elevation gains are less than surface accretion, then shallow subsidence of the soil volume is inferred (Extended Data Fig. 1). Conversely, if elevation gains are greater than surface accretion, then expansion of the subsurface soil profile is inferred, which may be due to root growth[8,9,15,23]. Over many sites it has been repeatedly shown that vertical accretion on the soil surface is not a valid substitute for surface elevation change and that the complete set-up is necessary to identify the contribution of surface and shallow subsurface processes to surface elevation change at a specific site[15,23,26]. Repeated measurements allow description of net surface elevation change, which can be integrated with region-specific relative SLR (for example, tide-gauge data; see Supplementary Information) to determine whether the surface elevation of mangroves has kept pace with SLR over that time period.

Analysis of variation in surface elevation. Linear regression was used to describe the relationships between: (1) surface elevation gains and accretion of sediment on the soil surface; (2) accretion on the surface and TSM; and (3) surface elevation gains and TSM. Forms of these relationships are given in the legend of Fig. 2.

The relative influence (in per cent) of predictor variables on surface elevation change (in millimetres per year) was analysed using boosted regression tree (BRT) models[29] developed using data from 153 observations and 10 predictors, with a tree complexity of 5 and learning rate of 0.005. We developed three models using three different measures of sea-level variation at each site (Supplementary Table 1): the long-term rate of SLR at tide gauges (model 1); the rates of sea-level change over the period of the surface elevation gain measurements at tide gauges (model 2); and the rates of sea-level change based on satellite altimetry (model 3). On the basis of cross-validation, the mean percentages (s.e.m.) of deviance explained by models 1–3 are 44.8% (±10.5%), 38.2% (±9.3%) and 40.9% (±8.9%), respectively. BRT modelling was done with *R* version 3.0.2 using packages *dismo* and *gbm*. The BRTs were built with a 10-fold cross-validation optimization, with a Gaussian distribution for surface elevation change. Stochasticity (bag fraction) was set to 0.5. The final models were fitted with 5,850 trees. Geomorphological setting followed the classifications of ref. 30: river delta, tidal, lagoon or carbonate island. Ecological habitat followed the classifications of ref. 31: fringe, scrub, hammock, basin, overwash or riverine. Dominant tree genera were *Avicennia*, *Rhizophora*, *Sonneratia*; mangrove forests were classified as mixed forests at sites where no single genera was dominant.

Estimating time to submergence. Years to submergence over variation in tidal range was estimated by summing annual elevation deficits (Extended Data Fig. 2). Elevation deficit is the difference between the rate of local SLR and the rates of surface elevation gain. Where elevation deficits in our data were observed ($N = 103$), mean elevation deficit over our sites was 6 mm yr^{-1} (dashed line in Extended Data Fig. 2). Extended Data Fig. 2 shows the years to submergence (on a logarithmic scale) of the highest intertidal mangrove forest over variation in tidal range (microtidal, blue; mesotidal, yellow; macrotidal, red), for a range of elevation deficits (1–20 mm yr^{-1}); see Extended Data Fig. 2.

Model to predict the year of submergence of mangrove ecosystems. A model to predict the year of submergence of mangrove ecosystems subject to accelerating rates of SLR was developed for various physical environmental contexts. The model was based on the observed rates of mangrove surface elevation change as a function of rate of SLR, suspended sediment availability and tidal range. The model was run from 2010 to 2100 in 10-year time steps (see Extended Data Fig. 3 for a summary of the modelling process).

The vertical range of mangrove distribution was assumed to be the upper 50% of the tidal range[22]. For example, if the tidal range was 1 m, the vertical distribution of mangroves was assumed to be 0.5 m. Assuming that mangroves were at their upper vertical limit of their range at the start of the simulations, the time until net

elevation loss was equivalent to 50% of the magnitude of tidal range (in metres) was calculated as the time to mangrove submergence. In each time step, elevation deficit was calculated as the magnitude of SLR minus the magnitude of surface elevation gain. Total elevation deficit over the 90-year simulation was calculated by summing the accumulated elevation deficits.

Elevation gain (in millimetres per year) caused by sediment accumulation for particular SLR and suspended-sediment scenarios was calculated according to the observed surface elevation data. The slope and intercept of linear models relating elevation gain to rate of SLR are given in Extended Data Table 2. The relationship between surface elevation gain and rate of SLR (in millimetres per year) was established for two TSM classes: low (<2.5 g m^{-3}) and high (>2.5 g m^{-3}). Linear regression was used to establish the functional forms of the relationships between surface elevation gain and the rate of SLR.

Scenarios of tidal ranges from 0.5 m to 2.0 m in 0.5-m increments were examined (4 total). Six SLR trajectories were simulated for a total of 24 simulations. The starting rate of SLR for each trajectory was 3 mm yr^{-1}, equivalent to the current rate of global average SLR. The rate of change of sea-level increase was varied by 0.5 mm yr^{-1} in decadal time steps for the six trajectories: SLR increased by 0.5 mm yr^{-1} each decade in the first trajectory, 1 mm yr^{-1} in the second, 1.5 mm yr^{-1} in the third and so on, up to 3.0 mm yr^{-1} each decade in the last trajectory. The resultant magnitudes of sea-level change for the six trajectories were 0.45 m, 0.63 m, 0.81 m, 0.99 m, 1.17 m and 1.35 m by 2100. The model was run for each of the tide-range ($N = 4$) and SLR ($N = 6$) trajectory combinations, for each of the sediment availability scenarios (low and high), for a total of 48 simulations.

We then created spatial layers of the model. The TSM layer was classified as high (>2.5 g m^{-3}) or low (<2.5 g m^{-3}). The tidal-range layer was sourced from the FES2012 tidal model package distributed by AVISO, with support from CNES (http://www.aviso.altimetry.fr/)[32]. We ran the model for each pixel that contains mangroves, as indicated by the data presented in ref. 4, for three SLR scenarios (RCP6.5, RCP8.5 and a higher, 1.4-m SLR by 2100 scenario based on ref. 33).

There are a number of assumptions and limitations to this approach. First, we assumed that mangroves commenced the simulations at the upper vertical limit of their range. Therefore, mangroves at lower vertical extents would submerge earlier and the model is an optimistic estimate of time until submergence. Second, while feedbacks between surface elevation change and other environmental features (sediment supply, vertical location in the tidal frame and so on) were not explicitly incorporated, they were implicitly included as they would have contributed to the observed SET data upon which the model was built. Third, we assumed that mangroves would be submerged when they reached MSL (50% of the tidal range). However, mangroves may be able to persist beyond this time (that is, there may be a time lag), owing to physiological tolerance and acclimatization. If this time lag were to exist, then it would extend the time frame for which mangroves would be expected to survive after submergence. Lastly, the model does not consider the area of habitat, or predict when new habitat would become available.

Time scales of soil surface elevation records. The timescale of SET measurements is relatively short compared to the timescales of ecosystem change in response to SLR; therefore, to assess whether SET measurements are representative of longer term rates of surface elevation change we took two approaches. The first uses SET records of differing lengths to compare shorter- and longer-term rates. The second compares SET elevation gains with those inferred from ^{210}Pb dating of sediment cores (over the scale of decades) for the few sites where sediment dating and SET data are available.

To assess whether the length of the SET record is likely to influence our results, surface elevation gains measured over longer periods (mean record length of 5.5 years) were compared to those over shorter periods (mean record length of 2.1 years) for three sites (New Zealand, $N = 3$; Micronesia, $N = 13$; Moreton Bay, Australia, $N = 18$). Longer-term and shorter-term rates were highly correlated ($R^2 = 0.59$) with a slope of 0.90 ± 0.13 which was not statistically different from 1 ($t = 0.769$, $P = 0.45$) (Extended Data Fig. 4). The lengths of the SET records were not correlated with surface elevation gain, surface accretion, shallow subsidence or elevation deficits relative to SLR. Six SETs in Micronesia have now been monitored for 16.6 years. At this site, surface elevation gains at 16.6 years were correlated with surface elevation gains at 6.6 years: (surface elevation at 16.6 yr) = $-0.16 + (0.36 \pm 0.12) \times$ (surface elevation at 6.6 yr), $R^2 = 0.59$. Thus, in Micronesia, the long-term elevation gain was approximately 40% of the short-term rate, indicating compaction of the sediment profile over time. [210]Pb dating of sediment cores and SET data are available for the New Zealand site and also for locations on the east coast of Australia. In New Zealand, sediment accumulation rates measured using SETs and those using [210]Pb (from the 1960s to the present) are similar[34]. In Moreton Bay, the mean rate of mangrove sediment accumulation using ^{210}Pb was 1.2 ± 0.9 mm yr^{-1}, which is lower, but in the range of that observed using SETs in similar habitats (1.7 ± 0.5 mm yr^{-1}; ref. 35); in southeastern Australia, ^{210}Pb was 1.7 ± 0.3 mm yr^{-1}, which is higher than that observed using SETs (0.72 ± 0.49 mm

yr^{-1}; ref. 35). Additionally, rates of surface elevation gain measured with SETs in the Caribbean and Florida are broadly consistent with sediment accumulation rates derived from ^{14}C dating[15] and ^{210}Pb dating[36]. The study in Florida[36] found sediment accumulation based on ^{210}Pb was on average 81% of that measured using 2.5-yr SET records. Compaction can be caused by loss of pore space due to dewatering and grain packing, and compression and decomposition of organic matter, which may not occur linearly over time[37]. Variation in sediment characteristics are likely to lead to variable rates of compaction over the Indo-Pacific region. If high rates of compaction are typical, then our short-term rates may over-estimate surface elevation gains for the region.

Total suspended matter in coastal waters. In this study we used level-3-processed TSM data at 4-km resolution and binned monthly (data freely available from http://hermes.acri.fr/). TSM concentration was derived from the MERIS instrument on the European Space Agency's (ESA) Envisat satellite (390–1,040 nm). TSM in coastal waters is an indicator of suspended sediments associated with river run-off and resuspension and is useful in both estuarine and reef lagoon waters[38]. Data products were processed and validated as part of the ESA's DUE GlobColour Global Ocean Colour for Carbon Cycle Research project (for more information on data processing see http://www.globcolour.info/CDR_Docs/GlobCOLOUR_PUG.pdf). The resulting raster grid was displayed in the plate-carrée projection. TSM was extracted using the open-source software BEAM VISAT (http://www.brockmann-consult.de/cms/web/beam/; ESA), using the TSM value of the pixel containing, or closest to, the SET site. For 24 sites, we used data from the pixel containing the SET site (that is, within 4 km of the site); 3 sites were 1 pixel distant and 3 sites were greater than 1 pixel distant (2 pixels for Kooragang Island, 5 pixels for Quail and 17 pixels for Porong). The Porong region has limited data availability owing to high cloud cover. An annual mean for 2011, where data from all sites was available, was calculated by averaging TSM pixel data for January, April, July and October. In other years, TSM from many sites were missing from the data set. We used mean annual data in 2011 to assess relationships between sediment accretion and TSM. Comparison of TSM values over the different years the data were available (since 2002) found that spatial differences were consistent over years. The relationship between mean TSM in 2011 and mean over the available record is shown in Extended Data Fig. 5. The linear regression of this relationship is (mean TSM in 2011) = $(1.38 \pm 0.94) + (0.58 \pm 0.08) \times$ (mean TSM over all available years), $R^2 = 0.64$, $P < 0.0001$, F test, where the indicated uncertainties are standard errors.

29. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77,** 802–813 (2008).
30. Woodroffe, C. in *Tropical Mangrove Ecosystems* (eds Robertson, A. I. & Alongi, D. M.) Ch. 2 (American Geophysical Union, 1993).
31. Lugo, A. I. & Snedaker, S. C. The ecology of mangroves. *Annu. Rev. Ecol. Syst.* **5,** 39–64 (1974).
32. Carrère, L., Lyard, F., Cancet, M., Guillot, A. & Roblou, L. FES 2012: a new global tidal model taking advantage of nearly 20 years of altimetry. In *Proc. 20 years of Progress in Radar Altimetry Symp.* ESA SP-710 (2012).
33. Horton, B. P., Rahmstorf, S., Engelhart, S. E. & Kemp, A. C. Expert assessment of sea-level rise by AD 2100 and AD 2300. *Quat. Sci. Rev.* **84,** 1–6 (2014).
34. Swales, A., Bentley, S. J. & Lovelock, C. E. Mangrove-forest evolution in a sediment-rich estuarine system: opportunists or agents of geomorphic change? *Earth Surf. Proc. Land.* **40,** 1672–1687 (2015).
35. Rogers, K., Saintilan, N. & Heijnis, H. Mangrove encroachment of salt marsh in Western Port Bay, Victoria: the role of sedimentation, subsidence and sea level rise. *Estuaries* **28,** 551–559 (2005).
36. Cahoon, D. R. & Lynch, J. C. Vertical accretion and shallow subsidence in a mangrove forest of southwestern Florida, U.S.A. *Mangroves Salt Marshes* **1,** 173–186 (1997).
37. Woodroffe, C. D. *et al.* Mangrove sedimentation and response to relative sea-level rise. *Annu. Rev. Mar. Sci.* Preprint at http://www.annualreviews.org/doi/abs/10.1146/annurev-marine-122414-034025.
38. Blondeau-Patissier, D. *et al.* ESA-MERIS 10-year mission reveals contrasting phytoplankton bloom dynamics in two tropical regions of Northern Australia. *Remote Sens.* **6,** 2963–2988 (2014).
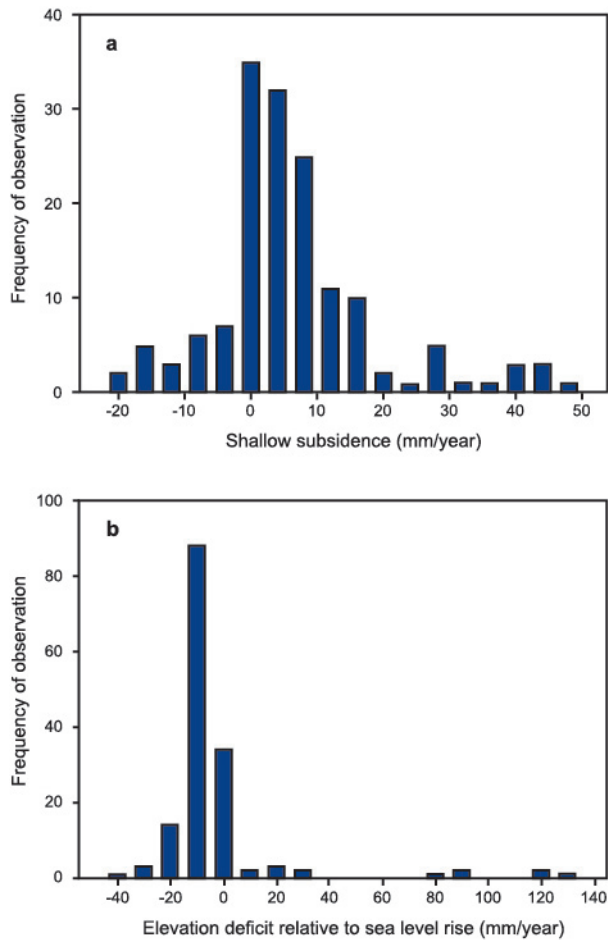
Extended Data Figure 1 | Frequency distributions of values of shallow subsidence and elevation deficits. a, The frequency distribution of shallow subsidence over all the SET sites, calculated as (surface accretion) − (surface elevation gain). (The data presented here are available online from the Source Data of Fig. 2). **b**, The frequency distribution of surface elevation deficits relative to SLR from tide gauges (see Supplementary Table 1).
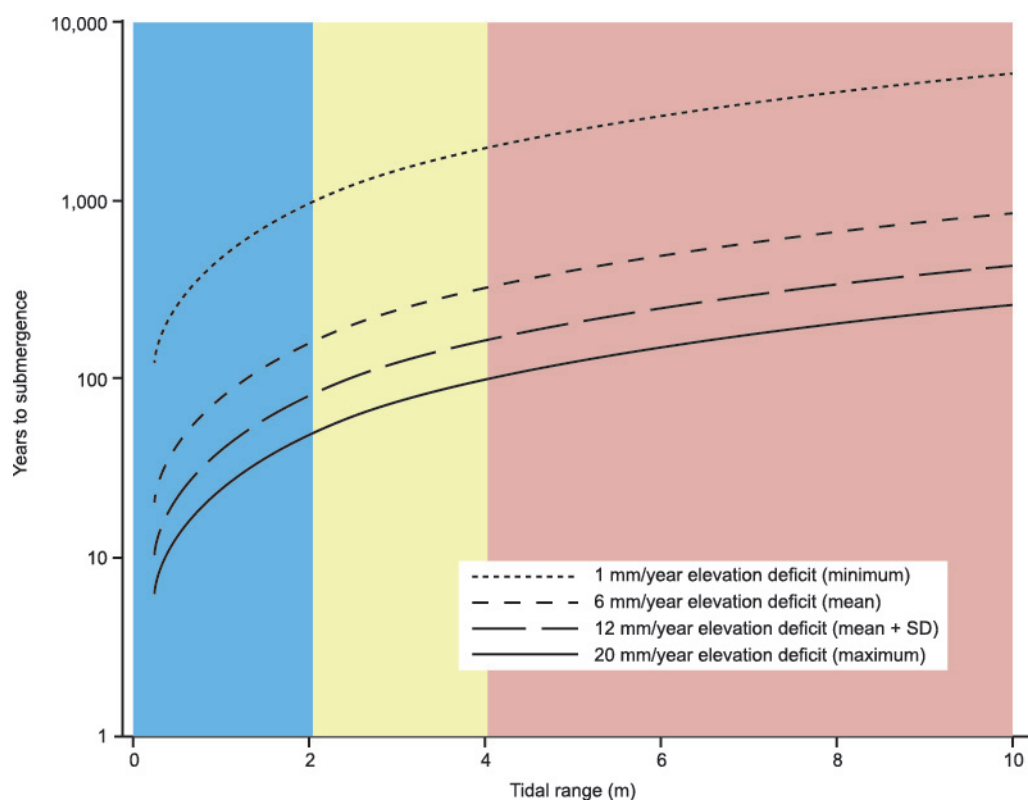
Extended Data Figure 2 | Years until submergence (logarithmic scale) of the highest intertidal mangrove forest over variation in tidal range and for a range of elevation deficits. The elevation deficit is the difference between the rate of local SLR and the rate of surface elevation gain. Submergence is assumed to occur when the cumulative elevation deficit is equivalent to the elevation capital (defined as half the tidal range). The mean elevation deficit in our study was $6 \, \text{mm yr}^{-1}$ (dashed line); other elevation deficits shown are $12 \, \text{mm yr}^{-1}$ (mean + SD = 6 + 6.3; long-dashed line), $1 \, \text{mm yr}^{-1}$ (minimum; dotted line) and $20 \, \text{mm yr}^{-1}$ (maximum; solid line). Categories of tidal range are coloured blue for microtidal, yellow for mesotidal and red for macrotidal.
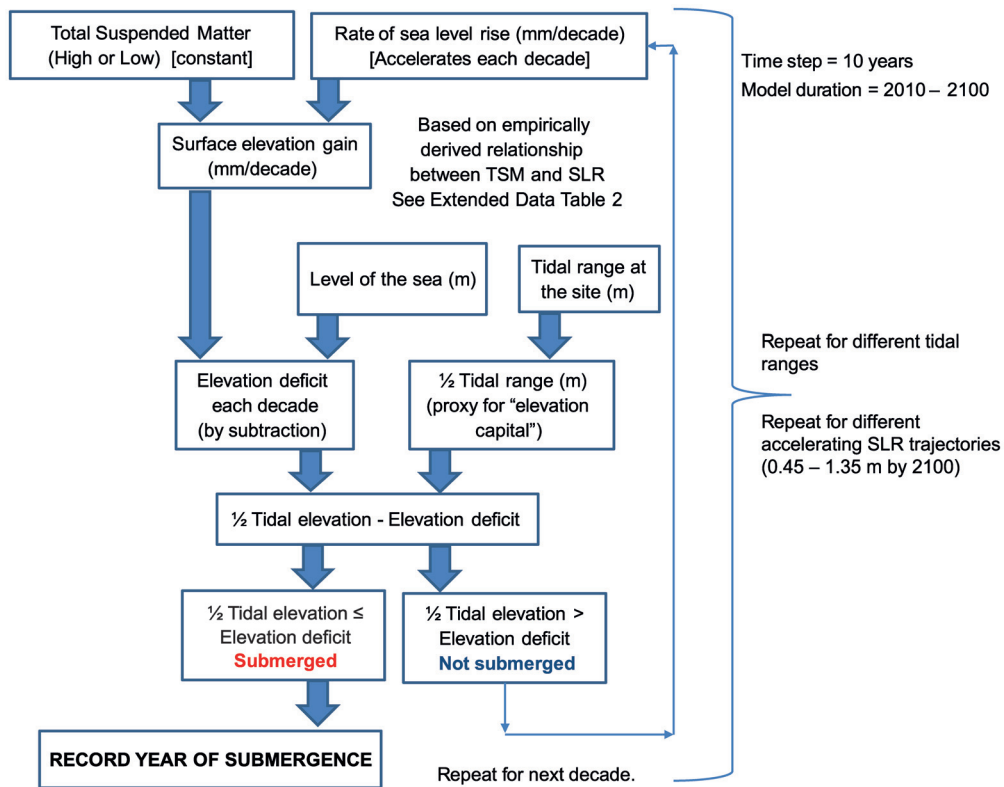
Extended Data Figure 3 | Schematic summary of the modelling process for estimating the decade of submergence of mangrove forests with SLR.
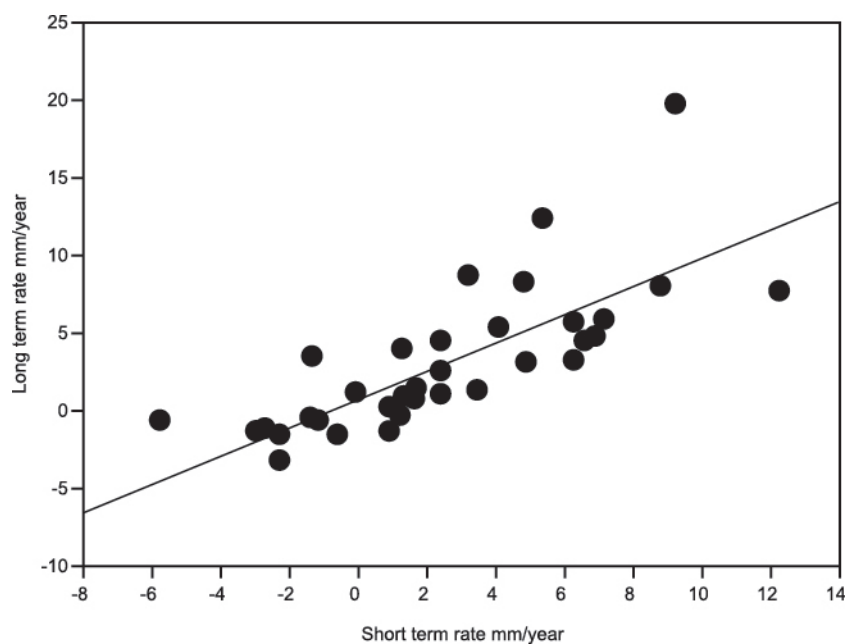
Extended Data Figure 4 | Comparison of surface elevation gains measured over longer and shorter periods for three sites. The three sites are New Zealand ($N = 3$), Micronesia ($N = 13$) and Moreton Bay, Australia ($N = 18$). The long-term mean record length is 5.5 years; the short-term mean record length is 2.1 years. Longer-term and shorter-term rates were highly correlated ($R^2 = 0.59$) with a slope of 0.90 ± 0.13, which is not statistically different from 1 ($t = 0.769$, $P = 0.45$).
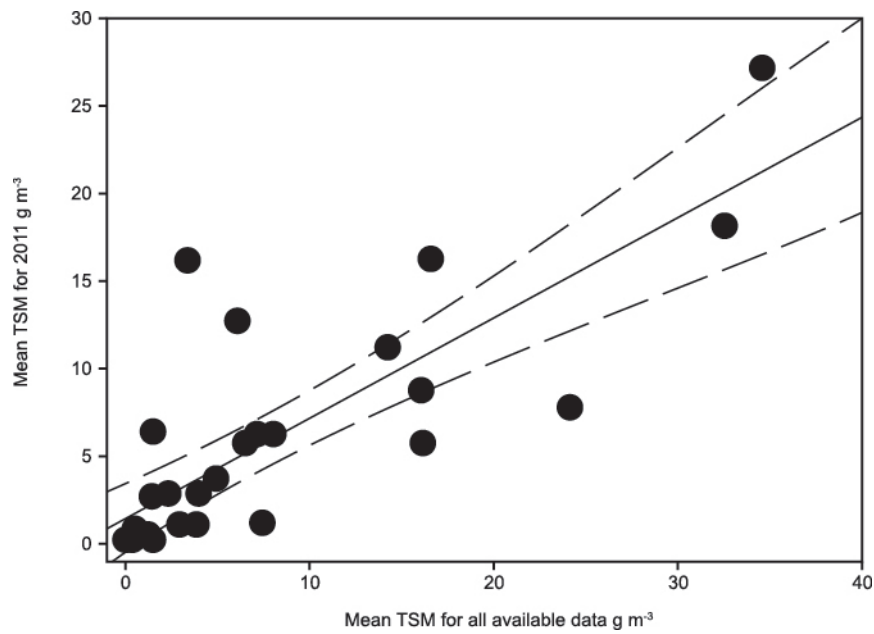
Extended Data Figure 5 | The relationship between mean TSM in 2011 over the available TSM record (2002–2011). In 2011, all sites were represented in the MERIS data set. The linear regression (solid line) of this relationship is (mean TSM in 2011) = (1.38 ± 0.94) + (0.58 ± 0.08) × (mean TSM over all available years), $R^2 = 0.64$, $P < 0.0001$, F test, where the indicated uncertainties are standard errors. Dashed lines are 95% confidence intervals.

Extended Data Table 1 | Summary of the relative influence of predictor variables on surface elevation change for BRT models

Model predictor	Relative Influence (%)
Model 1	
Total Suspended Matter (annual mean) g m^{-3}	36.47
Sea level change at tide gauge (mm/year)	29.32
Longitude	8.81
Geomorphological setting	7.19
Ecological habitat	5.36
Latitude	5.20
Dominant tree genera	3.68
Annual rainfall (mm)	2.84
Tidal range (m)	1.12
Model 2	
Total Suspended Matter (annual mean) g m^{-3}	37.40
Sea level change during the SET measurement (mm/year)	30.00
Longitude	8.56
Geomorphological setting	6.99
Ecological habitat	5.82
Latitude	4.00
Dominant tree genera	4.08
Annual rainfall (mm)	2.12
Tidal range (m)	1.06
Model 3	
Total Suspended Matter (annual mean) g m^{-3}	53.38
Sea level change from satellite altimetry (mm/year)	3.02
Longitude	14.22
Geomorphological setting	6.77
Ecological habitat	5.08
Latitude	4.97
Dominant tree genera	5.72
Annual rainfall (mm)	3.17
Tidal range (m)	3.66

See Methods for descriptions of the different models and predictors.

Extended Data Table 2 | Parameters used in the model for estimating time to submergence of mangrove forests for different sediment availability (classes of TSM).

Sediment availability classes	Slope	Standard error of estimate	Intercept	Standard error of estimate	R^2
Low (<2.5 g m^{-3})	0.36	0.04	0.29	0.53	0.45
High (>2.5 g m^{-3})	5.84	0.28	-24.6	2.49	0.85

The model is described in Extended Data Fig. 3. TSM values were obtained from MERIS data; see Methods. The parameters describe the linear regression relating surface elevation gain (in millimetres per year) to SLR (in millimetres per year) for two sediment availability bins, averaged over all tidal ranges; see Fig. 3.

LETTER

Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis

Minoru Takasato[1,2], Pei X. Er[1], Han S. Chiu[2], Barbara Maier[2], Gregory J. Baillie[2], Charles Ferguson[2], Robert G. Parton[2], Ernst J. Wolvetang[3], Matthias S. Roost[4], Susana M. Chuva de Sousa Lopes[4] & Melissa H. Little[1,2,5]

The human kidney contains up to 2 million epithelial nephrons responsible for blood filtration. Regenerating the kidney requires the induction of the more than 20 distinct cell types required for excretion and the regulation of pH, and electrolyte and fluid balance. We have previously described the simultaneous induction of progenitors for both collecting duct and nephrons via the directed differentiation of human pluripotent stem cells[1]. Paradoxically, although both are of intermediate mesoderm in origin, collecting duct and nephrons have distinct temporospatial origins. Here we identify the developmental mechanism regulating the preferential induction of collecting duct versus kidney mesenchyme progenitors. Using this knowledge, we have generated kidney organoids that contain nephrons associated with a collecting duct network surrounded by renal interstitium and endothelial cells. Within these organoids, individual nephrons segment into distal and proximal tubules, early loops of Henle, and glomeruli containing podocytes elaborating foot processes and undergoing vascularization. When transcription profiles of kidney organoids were compared to human fetal tissues, they showed highest congruence with first trimester human kidney. Furthermore, the proximal tubules endocytose dextran and differentially apoptose in response to cisplatin, a nephrotoxicant. Such kidney organoids represent powerful models of the human organ for future applications, including nephrotoxicity screening, disease modelling and as a source of cells for therapy.

The mammalian kidney is derived from intermediate mesoderm. Cells from the primitive streak (presomitic mesoderm; PSM) migrate rostrally to form the intermediate mesoderm[2]. The intermediate mesoderm gives rise to both key kidney progenitor populations, the ureteric epithelium and the metanephric mesenchyme, which form the collecting ducts and nephrons, respectively. Several studies have reported the successful differentiation of human pluripotent stem cells (hPSCs) into either ureteric epithelium or metanephric mesenchyme *in vitro*[3–7]. In contrast, we previously reported the simultaneous generation of both ureteric epithelium and metanephric mesenchyme from hPSCs, resulting in the induction of nephrons and collecting ducts[1]. This was paradoxical as it was assumed that the ureteric epithelium arises as a side branch of the mesonephric duct, itself forming from the anterior intermediate mesoderm, while the metanephric mesenchyme is derived from the posterior intermediate mesoderm[5,8]. Retinoic acid (RA) regulates anterior–posterior patterning in organogenesis with rostral RA signalling patterning the somites[9] (Fig. 1e). Conversely, the PSM expresses Cyp26, which attenuates RA signalling in the caudal embryo[10,11]. The PSM is also a strong site of Wnt signalling[12]. In our previous studies, we demonstrated *in vitro* that formation of the intermediate mesoderm required FGF9 or FGF2 (ref. 1). Hence, *in vivo* we assume that the ureteric epithelium forms from early migrating PSM cells exposed to FGF9 and RA soon after the primitive streak stage, while cells late to migrate, and hence exposed to longer Wnt signalling,

should give rise to the metanephric mesenchyme[13] (Fig. 1a). To confirm this, we varied the duration of initial Wnt signalling (using CHIR99021, an inhibitor of GSK-3) before addition of FGF9 (Fig. 1b) and monitored markers of the anterior intermediate mesoderm and posterior intermediate mesoderm by quantitative PCR. Shorter periods of CHIR99021 application induced the anterior intermediate mesoderm markers, *LHX1* and *GATA3*, whereas longer periods increased the posterior intermediate mesoderm markers, *HOXD11* and *EYA1*, at day 7. Prolonged expression of the PSM markers, *TBX6* and *T*, after a longer period in the presence of CHIR99021 suggested a delay in FGF9-induced fate commitment (Fig. 1c), as predicted. Immunofluorescence analysis showed that a longer (or shorter) period with CHIR99021 induced less (more) anterior intermediate mesoderm but more (less) posterior intermediate mesoderm, as indicated by GATA3 and HOXD11, respectively, at day 7 of differentiation (Fig. 1d). These observations persisted after 18 days of culture, with dominant ureteric epithelium induction (GATA3[+]PAX2[+]ECAD[+]) after fewer days in the presence of CHIR99021 and preferential induction of metanephric mesenchyme (PAX2[+]ECAD[−]) and its derivatives (PAX2[+]ECAD[+]) with more days in the presence of CHIR99021 (Extended Data Fig. 1a). Further, we investigated whether RA signalling also controls anterior–posterior fate patterning of the intermediate mesoderm using RA or an RA receptor antagonist, AGN193109, together with FGF9 (Fig. 1e, f). RA promoted ureteric epithelium induction, whereas AGN193109 inhibited ureteric epithelium but enhanced induction of the metanephric mesenchyme lineage (Fig. 1g and Extended Data Fig. 1b).

These results increase our understanding of embryogenesis as well as providing a method by which to modulate the relative induction of each of the two intermediate mesoderm-derived progenitor populations essential for kidney formation. As a result, we modified our existing kidney differentiation process to increase the proportion of metanephric mesenchyme formed, increase the time in 3D culture and actively trigger nephron formation. This optimized approach was applied to either human embryonic stem (ES) cells or human induced pluripotent stem (iPS) cells and involved an initial 4 days of CHIR99021, which resulted in the induction of both the ureteric epithelium and the metanephric mesenchyme in monolayer culture (Extended Data Fig. 2), followed by 3 days of FGF9 before transfer to organoid culture (Fig. 2a). The resulting aggregates were cultured for up to 20 days, during which time they spontaneously formed complex kidney organoids (Fig. 2b). During normal mouse kidney development, nephron formation from the metanephric mesenchyme is initiated in response to Wnt9b secreted from the ureteric epithelium. In the mouse, ectopic nephron formation can be triggered via the addition of canonical Wnt agonists[14]. Indeed, maximal nephron number per organoid required a pulse of CHIR99021 for one hour after forming a pellet (Fig. 2a and Extended Data Fig. 3a). In addition, the continued presence of FGF9 after this CHIR99021 pulse was essential

[1]Murdoch Childrens Research Institute, The Royal Children's Hospital Melbourne, Parkville, Victoria 3052, Australia. [2]Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia. [3]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St Lucia, Queensland 4072, Australia. [4]Department of Anatomy and Embryology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. [5]Department of Paediatrics, The University of Melbourne, Parkville, Victoria 3010, Australia.
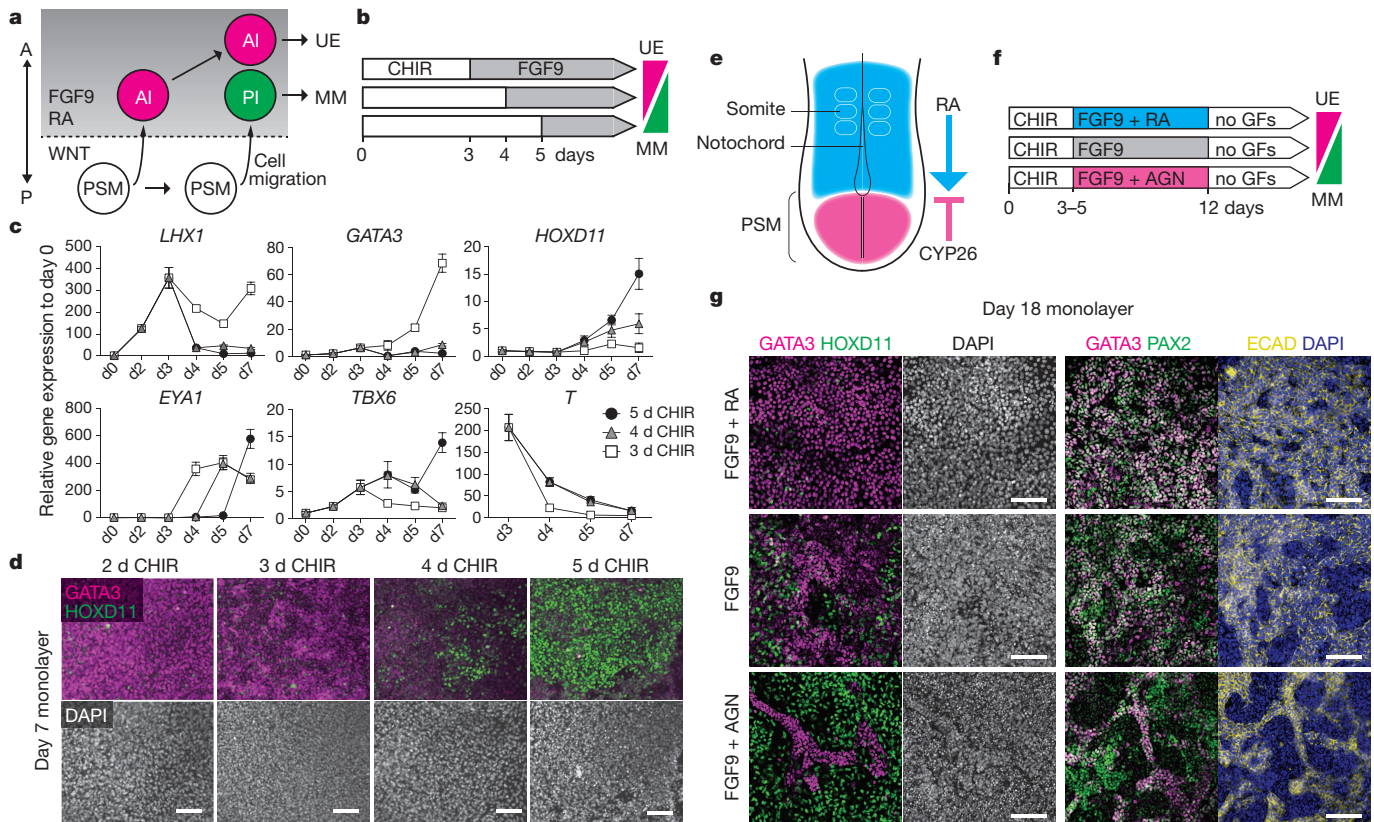
Figure 1 | Selective induction of either the collecting duct or kidney mesenchyme lineage. a, Schematic illustrating the mechanism of anterior–posterior (A–P) patterning of the intermediate mesoderm in the embryogenesis[13]. The timing of PSM cell migration determines the timing of the exposure to FGF9 and RA, resulting in fate selection between anterior intermediate mesoderm and posterior intermediate mesoderm. AI, anterior intermediate mesoderm; MM, metanephric mesenchyme; PI, posterior intermediate mesoderm; PSM, presomitic mesoderm; UE, ureteric epithelium. **b**, Schematic of three experimental timelines. CHIR, CHIR99021. **c**, Time course quantitative PCR of an initial 7 days (d) of the differentiation from the above timings. Experiments were conducted using monolayer culture condition (mean ± s.d., $n = 3$ independent experiments). **d**, Immunofluorescence at day 7 of differentiation with the AI marker, GATA3, and the PI

marker, HOXD11. Scale bars, 100 μm. Experimental replicates, 3. **e**, Schematic illustrating RA signalling after the primitive streak stage. An RA-metabolizing enzyme, CYP26, is expressed in the PSM region to shield PSM cells from RA signalling. **f**, Schematic of three experimental timelines. RA or AGN193109 (AGN) were added with FGF9 after CHIR99021, followed by growth factor withdrawal (no GFs). Experiments were conducted with monolayer culture condition. **g**, Immunofluorescence at day 18 of differentiation from 3 days of CHIR99021 followed by ± RA/AGN. AGN inhibited the AI specification of early migrating cells, causing posteriorization. At day 18, GATA3 and HOXD11 mark the UE and the MM, respectively (left panels). GATA3$^+$PAX2$^+$ECAD$^+$ cells represent the UE whereas GATA3$^-$ PAX2$^+$ cells do the MM (ECAD$^-$) and its derivatives (ECAD$^+$) (right panels). Experimental replicates, 3. Scale bars, 100 μm.

for nephrogenesis, suggesting an additional role for FGF signalling after Wnt-mediated nephron induction (Extended Data Fig. 3b). Within each organoid, the nephrons appropriately segmented into 4 components, including the collecting duct (GATA3$^+$ECAD$^+$), the early distal tubule (GATA3$^-$LTL$^-$ECAD$^+$), early proximal tubule (LTL$^+$ECAD$^-$) and the glomerulus (WT1$^+$) (Fig. 2c, d). Moreover, kidney organoids showed complex morphogenetic patterning with collecting duct trees forming at the bottom of the organoid, connecting to distal and proximal tubules in the middle, with the glomeruli at the top of each organoid (Fig. 2e and Supplementary Videos 1 and 2). This patterning mimics the tissue organization observed *in vivo* where glomeruli arise in the cortex whereas the collecting ducts radiate through the organ from the middle. Here again, the relative level of collecting duct versus nephron within individual organoids could be varied with the timing of the initial CHIR99021-to-FGF9 switch (Extended Data Fig. 4a, b). Next, we performed RNA sequencing of whole kidney organoids at day 0, 3, 11 and 18 after aggregation and 3D culture. Across this time course we observed a temporal loss of nephron progenitor gene expression but an increase in markers of multiple nephron segments, including the podocytes, proximal and distal tubules (Extended Data Fig. 5 and Supplementary Table 2). Transcriptional profiling was performed and compared using an unbiased method with human fetal transcriptional data sets from

21 human fetal organs/tissues from the first and/or second trimester of pregnancy[15]. This analysis clustered kidney organoids at d11 and d18 of culture with first trimester human fetal kidney (Fig. 2f, g and Extended Data Fig. 6). At the earlier culture time points (day 0 and 3), organoids more closely matched the fetal gonad, an embryologically closely related tissue also derived from the intermediate mesoderm.

In a kidney, the epithelial cell types (nephron and collecting duct) are surrounded by a renal interstitium (stroma) within which there is a vascular network. As well as forming the metanephric mesenchyme, the intermediate mesoderm gives rise to stromal and vascular progenitors (Fig. 3a)[16,17]. We examined kidney organoids for evidence of additional cell types and evidence of functional maturation. Collecting ducts could be distinguished based on co-expression of PAX2, GATA3 and ECAD (Fig. 3b). At d11, nephron epithelia showed proximal (LTL$^+$ECAD$^-$) and distal (LTL$^-$ECAD$^+$) elements (Fig. 3c). By day 18, proximal tubules matured to co-express LTL with ECAD, with cubilin evident on the apical surface (Fig. 3d, e). Transmission electron microscopy (TEM) showed distinct epithelial subtypes; cells with few short microvilli surrounding an open lumen characteristic of collecting duct/distal tubule (Fig. 3k) and typical proximal tubular epithelium displaying an apical brush border with tight junctions (Fig. 3l). By day 18, loops of Henle (UMOD$^+$) began to form (Fig. 3f). By day 11, WT1$^+$NPHS1$^+$ early glomeruli[18] comprising a Bowman's capsule
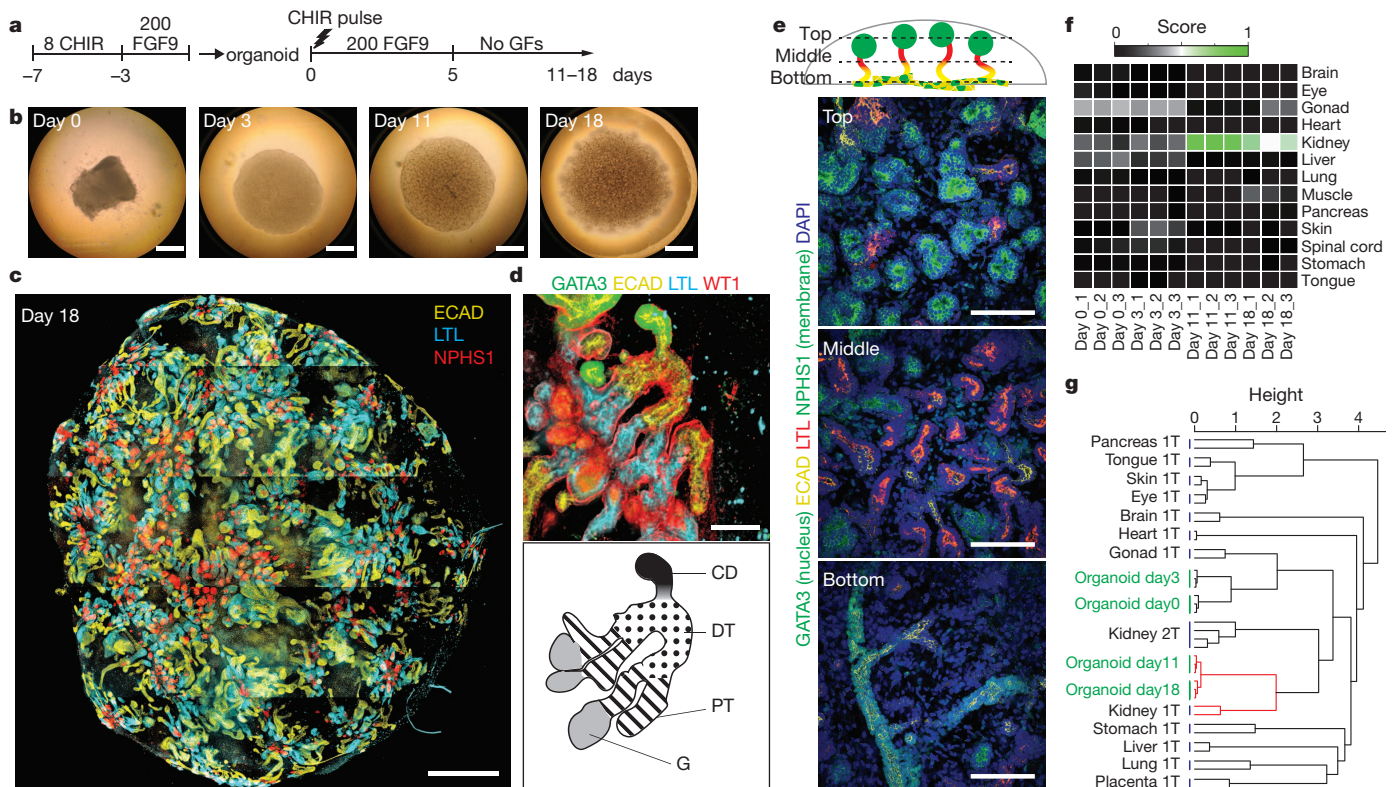
Figure 2 | Generating a kidney organoid equivalent to the human fetal kidney *in vitro*. a, Schematic of the differentiation protocol from hPSCs. 8 CHIR, 8 µM CHIR99021; 200 FGF9, 200 ng ml^{-1} FGF9. **b**, Global bright field observations of self-organizing kidney organoids across a time series. The success rate of organoid differentiation was 94.2% (138 organoids, 5 experiments). Scale bars, 1 mm. **c**, Tile scan immunofluorescence of a whole kidney organoid displaying structural complexity. Scale bar, 1 mm. **d**, High-power immunofluorescence microscopy showing a nephron segmented into 4 compartments, including the collecting duct (CD, GATA3$^+$ECAD$^+$), distal tubule (DT, GATA3$^-$ECAD$^+$LTL$^-$), proximal tubule (PT, ECAD$^-$LTL$^+$) and the glomerulus (G, WT1$^+$). Scale bar, 100 µm. **e**, Confocal microscopy generating serial *z*-stack images from the bottom to the top of a day 11 kidney organoid (Supplementary Videos 1 and 2). Schematic illustrates the position of different structures within an organoid. Top, middle and bottom images are representative images taken through the organoids at the position indicated in **e**. Each segment of the nephron is marked (or coloured in schematic) as described below: collecting ducts, GATA3$^+$ECAD$^+$ (green dots in yellow); distal tubules, ECAD$^+$ (yellow); proximal tubules, LTL$^+$ (red); glomeruli, NPHS1 (green circles). Scale bars, 100 µm. **f**, Heat map visualizing the relative transcriptional identity (score from 0 to 1 determined using the KeyGene algorithm[15]) of kidney organoids to 13 human fetal tissues. RNA-seq was performed on whole kidney organoids from 4 time points (day 0, 3, 11 and 18 after aggregation) with 3 individual organoids from 1 experiment per time point (see Supplementary Table 2). **g**, A dendrogram showing the hierarchical clustering of day 0, 3, 11 and 18 kidney organoids with human fetal organs from both first trimester and second trimester, based on 85 key genes (Supplementary Table 3) previously defined[15]. This clearly shows a close match with trimester 1 fetal kidney from day 11 and 18 of culture.

with central podocyte formation was seen connected to proximal tubules (Fig. 3g). Kidney organoids also developed a CD31$^+$ KDR$^+$SOX17$^+$ endothelial network with lumen formation (Fig. 3h and Extended Data Fig. 7a, b, c). TEM showed the presence of primary and secondary foot processes characteristic of podocytes (Fig. 3m). In a developing kidney, renal interstitium differentiates into pericytes and mesangial cells[19]. As expected, kidney organoids contained PDGFRA$^+$ perivascular cells that lie along KDR$^+$ endothelia and PDGFRA$^+$ early mesangial cells invaginating the glomeruli, as observed in human fetal kidney[20] (Extended Data Fig. 8a, b). Early avascular glomeruli contained basement membrane, as indicated by laminin staining and TEM, and showed attaching foot processes on the basement membrane (Extended Data Fig. 8c, d). In some instances, glomeruli showed evidence of endothelial invasion (Fig. 3i and Supplementary Videos 3 and 4), a feature never observed in explanted embryonic mouse kidneys[21]. Finally, nephrons were surrounded by MEIS1$^+$ renal interstitial cells[22], some of which were also FOXD1$^+$ (Fig. 3j and Extended Data Fig. 8e), suggesting the presence of cortical (FOXD1$^+$MEIS1$^+$) and medullary (FOXD1$^-$MEIS1$^+$) stroma. Hence, all anticipated kidney components form, pattern and begin to mature within these hPSC-derived kidney organoids. Consistent with these observations were the transcriptional changes across time in culture, with a gradual reduction in the nephrogenic mesenchyme and ureteric tip markers followed by the upregulation of genes specific to podocyte, proximal tubule, distal tubule and loop of Henle[23].(Extended Data Fig. 5).

The utility of stem-cell derived kidney organoids for disease modelling or drug screening will be dependent upon the functional maturation of the nephrons within these organoids. To test this, we focused on the proximal tubules, a nephron segment that has important roles in solute, vitamin, hormone and amino acids reabsorption. The capacity of cubilin-mediated proximal tubule specific endocytosis was demonstrated by the selective uptake of dextran–Alexa488 from the media by the LTL$^+$ tubules after 24 h of exposure (Fig. 4a and Extended Data Fig. 9a, b). The proximal tubules represent a particular target for nephrotoxicity due to the expression of multidrug resistance (such as ABCB1, ABCG2) and anion and cation transporters (such as the SLC22 gene family)[24]. Cisplatin is one such nephrotoxicant that induces caspase-mediated acute apoptosis of proximal tubular cells in the kidney[25,26]. We treated kidney organoids with 0, 5 and 20 µM cisplatin for 24 h before examining cleaved-CASP3 antibody staining (Extended Data Fig. 9c). While control organoids showed occasional apoptotic interstitial cells, both 5 µM and 20 µM cisplatin induced specific acute apoptosis in mature proximal tubular cells (LTL$^+$ECAD$^+$), whereas immature cells (LTL$^+$ECAD$^-$) did not undergo apoptosis (Fig. 4b, c).

In summary, this study demonstrates that by carefully balancing anterior–posterior patterning of intermediate mesoderm with small
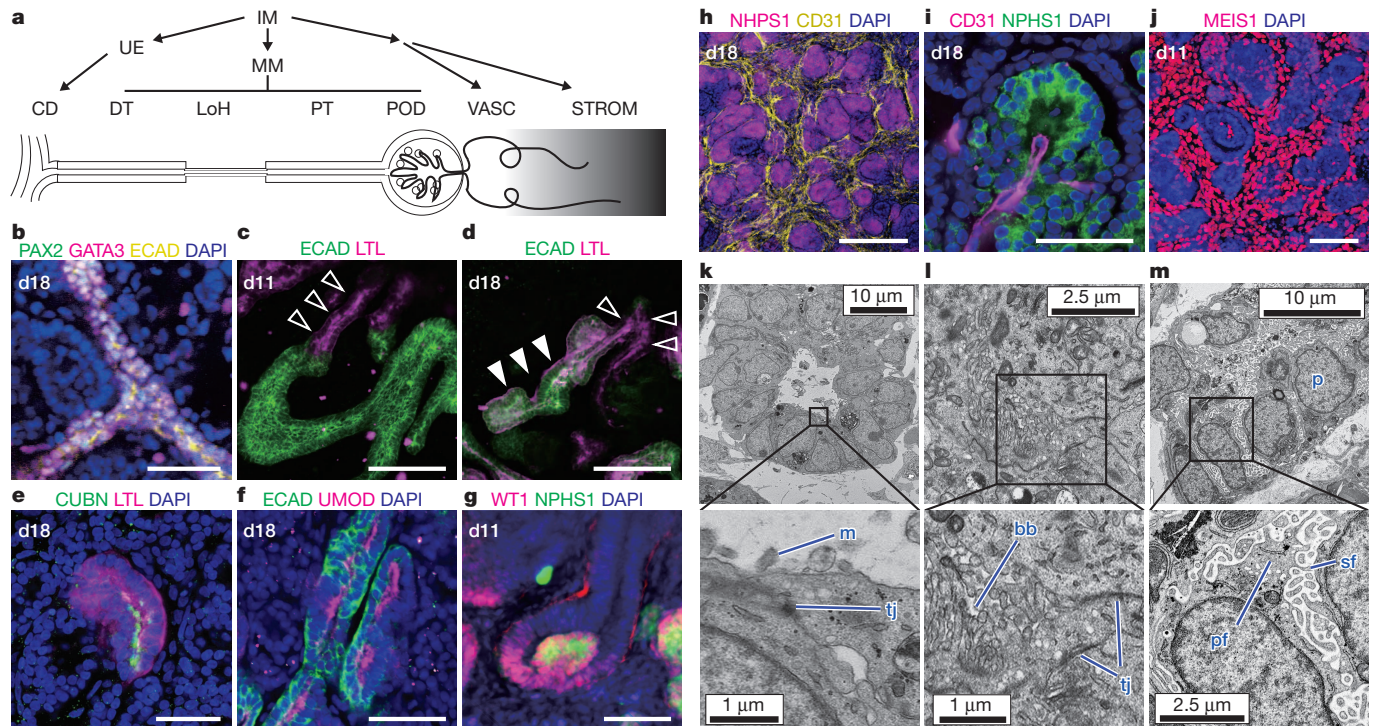
Figure 3 | Kidney organoids contain differentiating nephrons, stroma and vasculature with progressive maturation with time in culture. a, Schematic illustrating the developmental pathway from intermediate mesoderm (IM) to each cellular component of the kidney. CD, collecting ducts; DT, distal tubules; LoH, loops of Henle; PT, proximal tubules; POD, podocytes; VASC, vasculature; STROM, renal interstitium. **b–j**, Immunofluorescence of kidney organoids at either day 11 or 18. **b**, Collecting ducts marked by PAX2, GATA3 and ECAD. Scale bar, 50 μm. **c, d**, Early proximal tubules of LTL$^+$ECAD$^-$ at day 11 (black arrowheads). LTL$^+$ECAD$^+$ maturing proximal tubules appear by day 18 (white arrowheads). Scale bars, 100 μm. **e**, Proximal tubules express cubilin (CUBN). Scale bar, 50 μm. **f**, Loops of Henle marked by UMOD and ECAD.

Scale bar, 50 μm. **g**, A developing glomerulus with podocytes marked by WT1 and NPHS1. Scale bar, 50 μm. **h**, CD31$^+$ endothelia within the renal interstitium. Scale bar, 200 μm. **i**, Evidence of endothelial invasion into glomeruli at day 18 of culture. Scale bar, 50 μm. **j**, The kidney interstitium marked by MEIS1. Scale bar, 100 μm. **k–m**, Transmission electron microscopy of kidney organoids. **k**, A putative distal tubule with relatively sparse short microvilli (m) and tight junctions (tj). **l**, A putative proximal tubule with a lumen filled with extensive closely packed microvilli characteristic of the brush border (bb). **m**, Podocytes (p) with characteristic large nuclei and primary (pf) and secondary foot (sf) processes. Data are representative from a minimum of 3 independent experiments.

molecules it is possible to direct human pluripotent stem cells to form a complex multicellular kidney organoid that comprises fully segmented nephrons surrounded by endothelia and renal interstitium and is transcriptionally similar to a human fetal kidney. As such, these will improve our understanding of human kidney development. Each kidney organoid reaches a substantial size with more than 500 nephrons per organoid, a number equivalent to a mouse kidney at 14.5 days post-coitum[27]. While there is room for further improvement with regard to
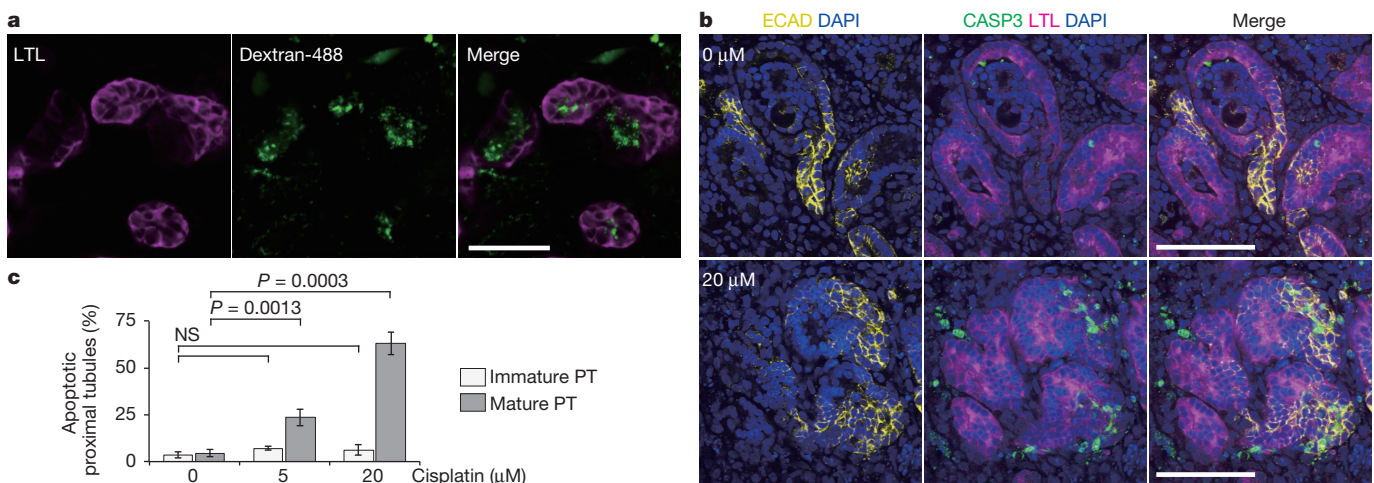
Figure 4 | Functional maturation of the proximal tubule. a, Dextran uptake assay showing endocytic ability of LTL$^+$ tubules. Scale bar, 50 μm. **b**, Treating kidney organoids with 20 μM cisplatin caused apoptosis in LTL$^+$ECAD$^+$ proximal tubular cells. Apoptotic cells were detected by cleaved caspase 3 antibody-staining (CASP3). Scale bars, 100 μm. **c**, Quantification of the number of apoptotic tubules showing mature proximal tubules-specific apoptosis by a nephrotoxicant, cisplatin. In response to 5 μM and 20 μM cisplatin, LTL$^+$ECAD$^+$ mature proximal tubules (PT) underwent apoptosis dose-dependently. In contrast, LTL$^+$ECAD$^-$ immature PT did not respond to cisplatin. P values were calculated by independent t-test (mean ± s.e.m., $n = 5$ independent experiments); NS, not significant.

tubular functional maturity, glomerular vascularisation and a contiguous collecting duct epithelium with a single exit path for urine, the tissue complexity and degree of organoid functionalization observed here supports their use to screen drugs for toxicity, modelling genetic kidney disease or act as a source of specific kidney cell types for cellular therapy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Takasato, M. *et al.* Directing human embryonic stem cell differentiation towards a renal lineage generates a self-organizing kidney. *Nature Cell Biol.* **16,** 118–126 (2014).
2. James, R. G. & Schultheiss, T. M. Patterning of the avian intermediate mesoderm by lateral plate and axial tissues. *Dev. Biol.* **253,** 109–124 (2003).
3. Mae, S. *et al.* Monitoring and robust induction of nephrogenic intermediate mesoderm from human pluripotent stem cells. *Nat. Commun.* **4,** 1367 (2013).
4. Xia, Y. *et al.* Directed differentiation of human pluripotent cells to ureteric bud kidney progenitor-like cells. *Nature Cell Biol.* **15,** 1507–1515 (2013).
5. Taguchi, A. *et al.* Redefining the *in vivo* origin of metanephric nephron progenitors enables generation of complex kidney structures from pluripotent stem cells. *Cell Stem Cell* **14,** 53–67 (2014).
6. Lam, A. Q. *et al.* Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. *J. Am. Soc. Nephrol.* **25,** 1211–1225 (2014).
7. Kang, M. & Han, Y. M. Differentiation of human pluripotent stem cells into nephron progenitor cells in a serum and feeder free system. *PLoS One* **9,** e94888 (2014).
8. Xu, J. *et al.* Eya1 interacts with Six2 and Myc to regulate expansion of the nephron progenitor pool during nephrogenesis. *Dev. Cell* **31,** 434–447 (2014).
9. Duester, G. Retinoic acid synthesis and signaling during early organogenesis. *Cell* **134,** 921–931 (2008).
10. Sakai, Y. *et al.* The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. *Genes Dev.* **15,** 213–225 (2001).
11. Abu-Abed, S. *et al.* The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15,** 226–240 (2001).
12. Sweetman, D., Wagstaff, L., Cooper, O., Weijer, C. & Münsterberg, A. The migration of paraxial and lateral plate mesoderm cells emerging from the late primitive streak is controlled by different Wnt signals. *BMC Dev. Biol.* **8,** 63 (2008).
13. Takasato, M. & Little, M. H. The origin of the mammalian kidney: implications for recreating the kidney *in vitro*. *Development* **142,** 1937–1947 (2015).
14. Park, J. S. *et al.* Six2 and Wnt regulate self-renewal and commitment of nephron progenitors through shared gene regulatory networks. *Dev. Cell* **23,** 637–651 (2012).
15. Roost, M. S. *et al.* KeyGenes, a tool to probe tissue differentiation using a human fetal transcriptional atlas. *Stem Cell Reports* **4,** 1112–1124 (2015).
16. Mugford, J. W., Sipilä, P., McMahon, J. A. & McMahon, A. P. Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney. *Dev. Biol.* **324,** 88–98 (2008).
17. Sims-Lucas, S. *et al.* Endothelial progenitors exist within the kidney and lung mesenchyme. *PLoS One* **8,** e65993 (2013).
18. Brunskill, E. W., Georgas, K., Rumballe, B., Little, M. H. & Potter, S. S. Defining the molecular character of the developing and adult kidney podocyte. *PLoS One* **6,** e24640 (2011).
19. Kobayashi, A. *et al.* Identification of a multipotent self-renewing stromal progenitor population during mammalian kidney organogenesis. *Stem Cell Reports* **3,** 650–662 (2014).
20. Floege, J. *et al.* Localization of PDGF alpha-receptor in the developing and mature human kidney. *Kidney Int.* **51,** 1140–1150 (1997).
21. Loughna, S., Yuan, H. T. & Woolf, A. S. Effects of oxygen on vascular patterning in Tie1/LacZ metanephric kidneys *in vitro*. *Biochem. Biophys. Res. Commun.* **247,** 361–366 (1998).
22. Brunskill, E. W. *et al.* Atlas of gene expression in the developing kidney at microanatomic resolution. *Dev. Cell* **15,** 781–791 (2008).
23. Thiagarajan, R. D. *et al.* Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PLoS One* **6,** e17286 (2011).
24. Cheng, X. & Klaassen, C. D. Tissue distribution, ontogeny, and hormonal regulation of xenobiotic transporters in mouse kidneys. *Drug Metab. Dispos.* **37,** 2178–2185 (2009).
25. Mese, H., Sasaki, A., Nakayama, S., Alcalde, R. E. & Matsumura, T. The role of caspase family protease, caspase-3 on cisplatin-induced apoptosis in cisplatin-resistant A431 cell line. *Cancer Chemother. Pharmacol.* **46,** 241–245 (2000).
26. Cummings, B. S. & Schnellmann, R. G. Cisplatin-induced renal cell apoptosis: caspase 3-dependent and -independent pathways. *J. Pharmacol. Exp. Ther.* **302,** 8–17 (2002).
27. Short, K. M. *et al.* Global quantification of tissue dynamics in the developing mouse kidney. *Dev. Cell* **29,** 188–202 (2014).

Supplementary Information is available in the online version of the paper.

METHODS

No statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture and differentiation. All experiments presented used the previously described wild-type human iPS cell line CRL1502 (clone C32) generated using episomal reprogramming[28]. Undifferentiated human iPS cells were maintained on the mouse embryonic fibroblasts (MEFs) (Millipore) as a feeder layer with human ES cell (hES) medium as described previously[1]. Cells were authenticated and tested for the mycoplasma infection[28]. Human iPS cells were plated on a Matrigel-coated (Millipore) culture dish and cultured in MEF-conditioned hES medium (MEF-CM) until reaching 60–100% confluence. Then, cells were again plated on a Matrigel-coated at 5,000 cells per cm^2 in MEF-CM. Next day, cells reached 40–50% of confluence, cells were treated with 8 μM CHIR99021 in APEL basal medium (STEMCELL Technologies) supplemented with Antibiotic-Antimycotic (Life Technologies) for 2–5 days, followed by FGF9 (200 ng ml^{-1}) and heparin (1 μg ml^{-1}) for another 5–2 days, with changing medium every second day. At day 7, cells were collected and dissociated into single cells using trypsin or TrypLE select (Life Technologies). Cells (0.5 × 10^6) were spun down at ×400g for 2 min to form a pellet and then transferred onto a Transwell 0.4 μm pore polyester membrane (CLS3450 Corning). Pellets were treated with 5 μM CHIR99021 in APEL for 1 h, and then cultured with FGF9 (200 ng ml^{-1}) and heparin (1 μg ml^{-1}) for 5 days, followed by another 6–13 days in APEL basal medium, with changing medium three times a week. Culture medium should not overflow over the membrane. For the differentiation in monolayer cultures, cells after CHIR99021 induction were treated by FGF9 (200 ng ml^{-1}) and heparin (1 μg ml^{-1}) for 10 days, followed by APEL basal medium for another 6 days. In some experiments, RA (0.1 μM) or AGN193109 (5 μM) were added to FGF9 medium. A step-by step protocol describing kidney organoid generation can be found at Protocol Exchange[29].

Immunocytochemistry. For monolayer cells, antibody staining was performed as described previously[1]. For the kidney organoid, organoids were fixed with 2% paraformaldehyde in PBS for 20 min at 4 °C followed by 3 times wash with PBS. Then organoids were blocked with 10% donkey serum, 0.3% Triton X/PBS for 2–3 h at room temperature and incubated with primary antibodies overnight at 4 °C. After 5 times washing with 0.1% Triton X/PBS, secondary antibodies were incubated for 4 h at room temperature. The following antibodies and dilutions were used: rabbit anti-PAX2 (1:300, 71-6,000, Zymed Laboratories), goat anti-SIX1 (1:300, sc-9709, Santa Cruz Biotechnology), rabbit anti-SIX2 (1:300, 11562-1-AP, Proteintech), mouse anti-ECAD (1:300, 610181, BD Biosciences), rabbit anti-WT1 (1:100, sc-192, Santa Cruz Biotechnology), mouse anti-HOXD11 (1:300, SAB1403944, Sigma-Aldrich), goat anti-GATA3 (1:300, AF2605, R&D Systems), rabbit anti-JAG1 (1:300, ab7771, Abcam), goat anti-cubilin (1:150, sc-20607, Santa Cruz Biotechnology), sheep anti-NPHS1 (1:300, AF4269, R&D Systems), LTL-biotin-conjugated (1:300, B-1325, Vector Laboratories), DBA-biotin-conjugated (1:300, B-1035, Vector Laboratories), mouse anti-KRT8 (1:300, TROMA, DSHB), mouse anti-CD31 (1:300, 555444, BD Pharmingen), rabbit anti-KDR (1:300, 2479, Cell Signaling Technology), goat anti-SOX17 (1:300, AF1924, R&D Systems), mouse anti-PDGFRA (1:200, 556001, BD Pharmingen), rabbit anti-Laminin (1:300, L9393, Sigma-Aldrich), rabbit anti-UMOD (1:300, BT-590, Biomedical Technologies), mouse anti-MEIS1 (1:300, ATM39795, activemotif), goat anti-FOXD1 (1:200, sc-47585, Santa Cruz Biotechnology) and rabbit anti-cleaved-CASP3 (1:300, 9661, Cell Signaling Technology). Images were taken using a Nikon Ti-U microscope or a Zeiss LSM 780 confocal microscope. All immunofluorescence analyses were successfully repeated more than three times and representative images are shown.

Electron microscopy. Organoids were processed for electron microscopy using a method as follows. A solution of 5% glutaraldehyde in 2 × PBS was added directly to the organoid culture dish in equal volume to the growth medium and placed under vacuum for 5 min. The organoid was reduced in size by cutting into small blocks (~2 × 2 mm), and irradiated in fresh fixative 2.5%, again under vacuum, for 6 min, in a Pelco Biowave (Ted Pella In, Redding, CA) at 80 W power. Samples were then washed 4 × 2 min in 0.1 M cacodylate buffer. Samples were then immersed in a solution containing potassium ferricyanide (3%) and osmium tetroxide (2%) in 0.1 M cacodylate buffer for 30 min at room temperature. Following 6 × 3 min washes in distilled water the tissue blocks were then incubated in a filtered solution containing thiocarbohydrazide (1%) for 30 min at room temperature. After subsequent washing in distilled water (6 × 2 min) samples were incubated in an aqueous solution of osmium tetroxide (2%) for 30 min, then in distilled water (6 × 2 min) and incubated in 1% aqueous uranyl acetate for 30 min at 4 °C. After further distilled water washes (2 × 2 min) a freshly prepared filtered solution of 0.06% lead nitrate in aspartic acid (pH 5.5) warmed to 60 °C was added to the dish and further incubated for 20 min at 60 °C before rinsing in distilled water (6 × 3 min) at room temperature. Tissue blocks were dehydrated twice in each ethanol solution of 30%, 50%, 70%, 90% and absolute ethanol for 40 s at 250 W in the Pelco Biowave. Epon LX112 resin was used for embedding the tissue with infiltration at 25%, 50%, and 75% resin:absolute ethanol in the Pelco Biowave under vacuum at 250 W for 3 min and finishing with 100% resin (twice), before the final embedding/blocking and curing at 60 °C for 12 h.

qRT–PCR analysis. Total RNA was extracted from cells using Purelink RNA mini kit (Life Technologies) and cDNA was synthesized from >100 ng total RNA using Super Script III reverse transcriptase (Life Technologies). qRT–PCR analyses were performed with GoTaq qPCR Master Mix (Promega) by Roche LightCycler 96 real-time PCR machine. All absolute data were first normalized to GAPDH and then normalized to control samples ($\Delta\Delta C_t$ method). The sequences of primers used for qRT–PCR are as listed in Supplementary Table 1.

Next generation RNA sequencing and comparative analysis using KeyGenes. Sequencing was performed using the Illumina NextSeq500 (NextSeq control software v1.2/Real Time Analysis v2.1) platform. The library pool was diluted and denatured according to the standard NextSeq500 protocol and sequencing was carried out to generate single-end 76 bp reads using a 75 cycle NextSeq500 High Output reagent Kit (Catalog FC-404-1005). Reads were mapped against the reference human genome (hg19) using STAR[30], and read counts for each gene in the UCSC annotation were generated using htseq-count in the HTSeq python package (http://www-huber.embl.de/users/anders/HTSeq/doc/index.html). The number of uniquely mapped reads ranged from 18,810,634 to 36,706,805 per sample. Normalized read counts were calculated using the DESeq2 package[31].

KeyGenes was used to generate the identity scores of day 0, 3, 11 and 18 kidney organoids to different first trimester human organs, including the kidneys (GSE66302)[15]. The dendrogram showing the hierarchical clustering of day 0, 3, 11 and 18 kidney organoids and 21 human fetal organs from first and second trimester (GSE66302) was based on the Pearson correlation of the expression levels of 85 classifier genes as determined by KeyGenes (http://www.keygenes.nl) (Supplementary Table 3). The classifier genes were calculated by KeyGenes using the top 500 most differentially expressed genes of the human fetal data without including the extraembryonic tissues from that data set.

Functional analysis for proximal tubules. For dextran uptake assay, organoids at day 17 were cultured with 10 μg ml^{-1} of 10,000 MW dextran Alexa488-conjugated (D-22910, Life Technologies) for 24 h. Organoids were fixed and stained by LTL without permeabilization. For nephrotoxicity assays, organoids at day 17 were cultured with 0, 5, 20 or 100 μM cisplatin (Sigma-Aldrich) for 24 h. The ratio of apoptotic proximal tubules to total proximal tubules was manually counted using ImageJ in 2 or 3 representative fields per experiment. In total, $n = 5$ independent experiments. Images were taken using Zeiss LSM 780 confocal microscope.

28. Briggs, J. A. et al. Integration-free induced pluripotent stem cells model genetic and neural developmental features of down syndrome etiology. *Stem Cells* **31,** 467–478 (2013).
29. Takasato, M., Er, X. P., Chiu, S. H. & Little, H. M. Generation of kidney organoids from human pluripotent stem cells. *Protoc. Exch.* http://dx.doi.org/10.1038/protex.2015.087 (2015).
30. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
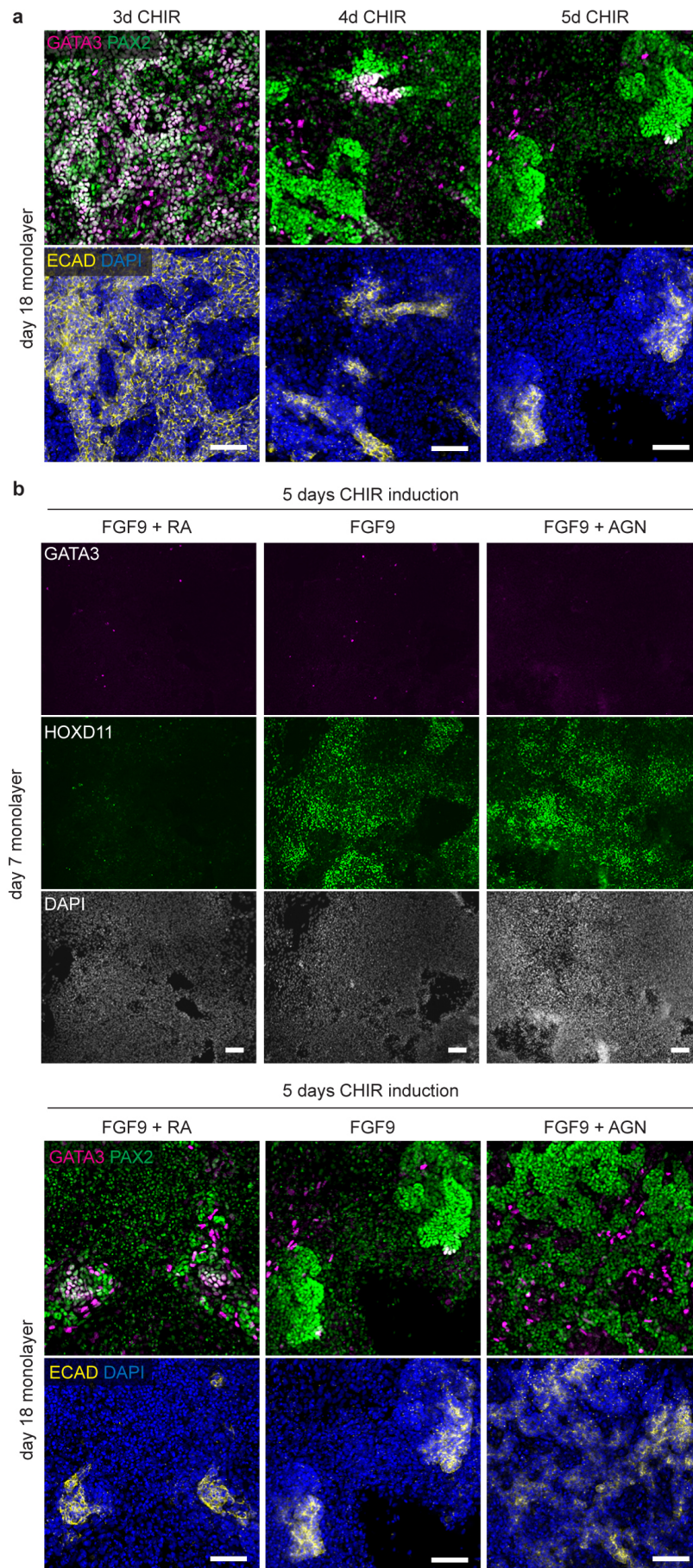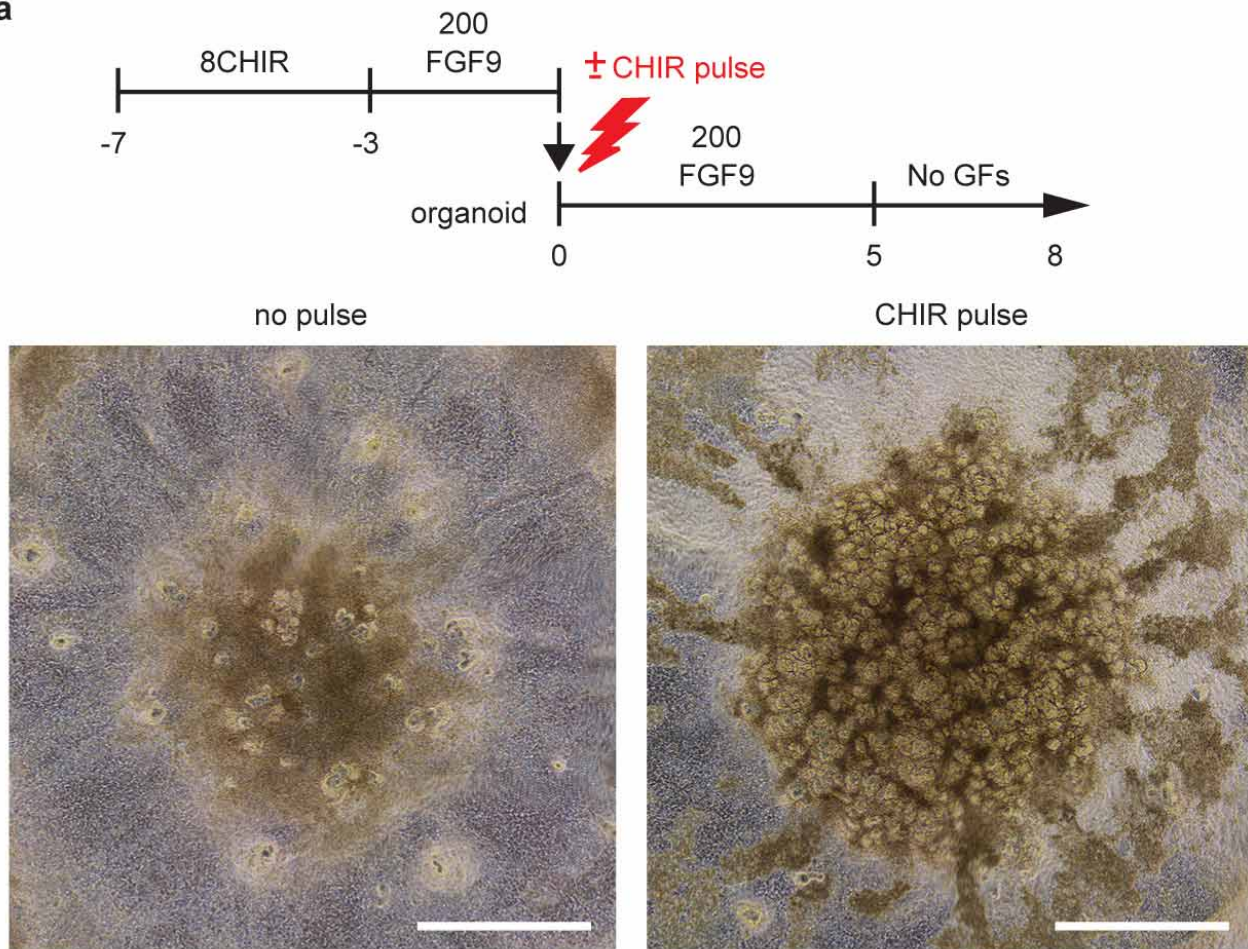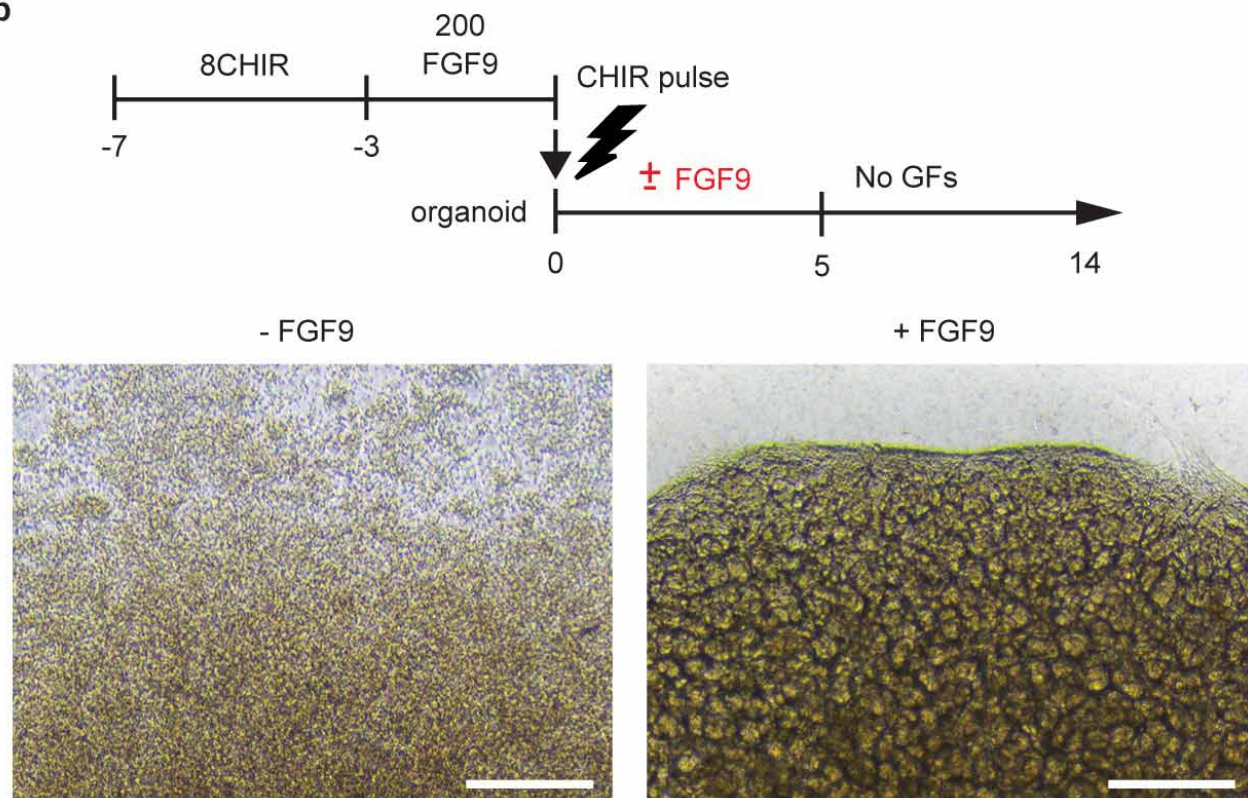
Extended Data Figure 1 | Antero-posterior intermediate mesoderm specification is regulated by the timing of FGF9 exposure and the presence of RA signalling. a, Immunofluorescence at day 18 of monolayer differentiation from cultures exposed to different timing of FGF9 addition (after 2, 3, 4 and 5 days of CHIR99021). The ureteric epithelium is represented by GATA3$^+$PAX2$^+$ECAD$^+$ cells. The metanephric mesenchyme and its derivatives are marked by PAX2$^+$GATA3$^-$ECAD$^-$ (metanephric mesenchyme) and PAX2$^+$GATA3$^-$ECAD$^+$ (nephrons), respectively. Scale bars, 100 μm. **b**, Immunofluorescence at day 7 and 18 of monolayer differentiation using 5 days of CHIR99021 followed by RA or AGN193109 (AGN) on top of FGF9. RA reduced the specification of posterior intermediate mesoderm, as indicated by the reduction of HOXD11 at day 7 (top panel). This resulted in less metanephric mesenchyme but some ureteric epithelium by RA at day 18 (bottom panel). Scale bars, 100 μm.

a

b
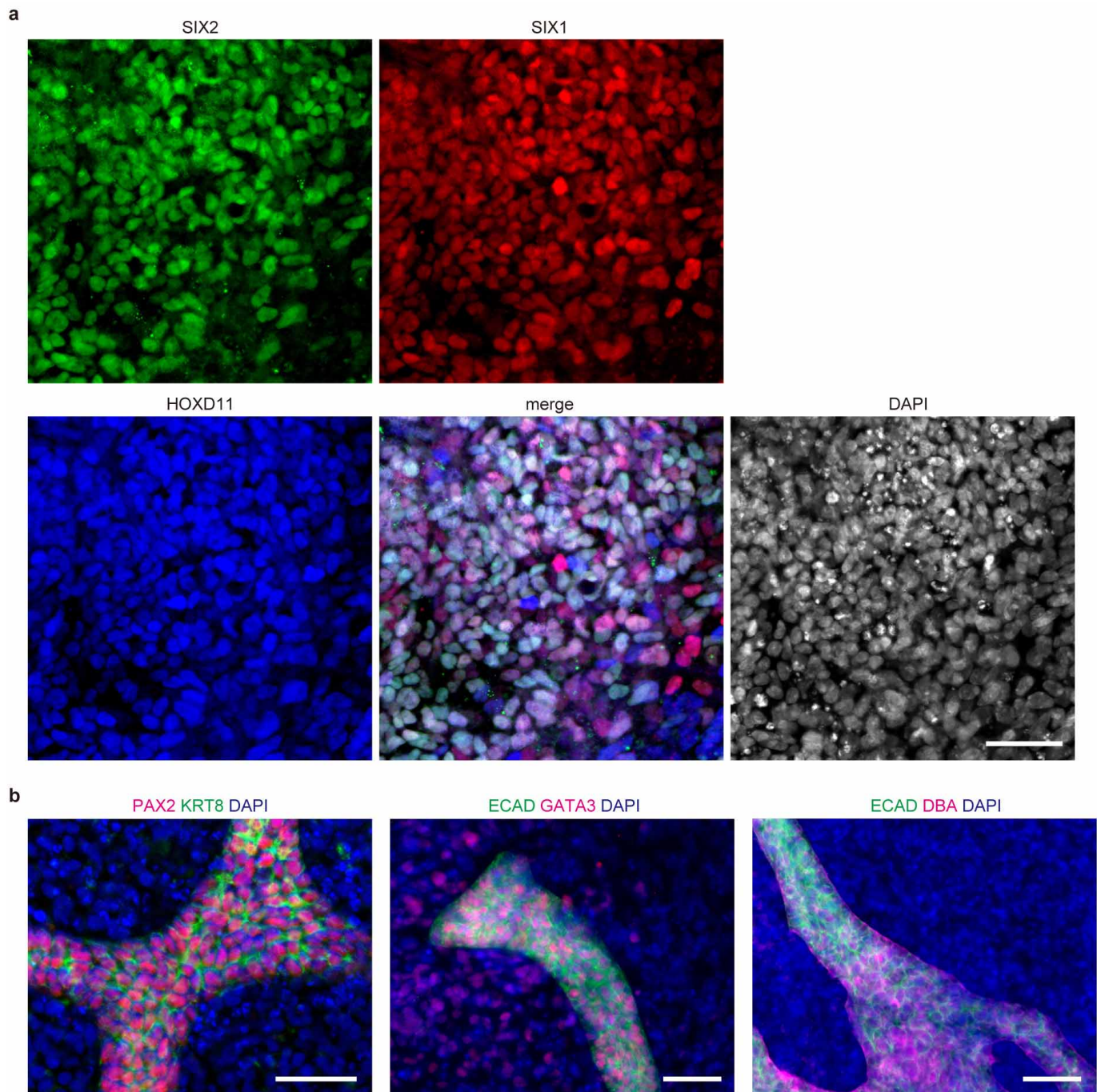
Extended Data Figure 2 | Induction of both kidney progenitors at the same time. a, b, Immunofluorescence at day 18 of the monolayer differentiation using the 4 days CHIR99021 before FGF9 protocol. The metanephric mesenchyme is marked by SIX2$^+$SIX1$^+$HOXD11$^+$ cells (**a**). GATA3$^+$ PAX2$^+$ECAD$^+$KRT8$^+$ cells representing the ureteric epithelium were also induced (**b**). Scale bars, 50 μm.

a

8CHIR · 200 FGF9 · ± CHIR pulse

-7 · -3 · 200 FGF9 · No GFs

organoid · 0 · 5 · 8

no pulse · CHIR pulse

b

8CHIR · 200 FGF9 · CHIR pulse

-7 · -3 · ± FGF9 · No GFs

organoid · 0 · 5 · 14

- FGF9 · + FGF9

Extended Data Figure 3 | Regulation of nephrogenesis in the kidney organoid. a, Stimulating organoids with 5 μM CHIR99021 for 1 h immediately after aggregation promoted nephrogenesis (CHIR pulse), whereas only limited numbers of nephrogenesis events happened without CHIR99021 (no pulse). Scale bars, 1 mm. **b**, Without the addition of FGF9 after this CHIR99021 pulse, organoids did not initiate nephrogenesis (−FGF9). Scale bars, 200 μm.
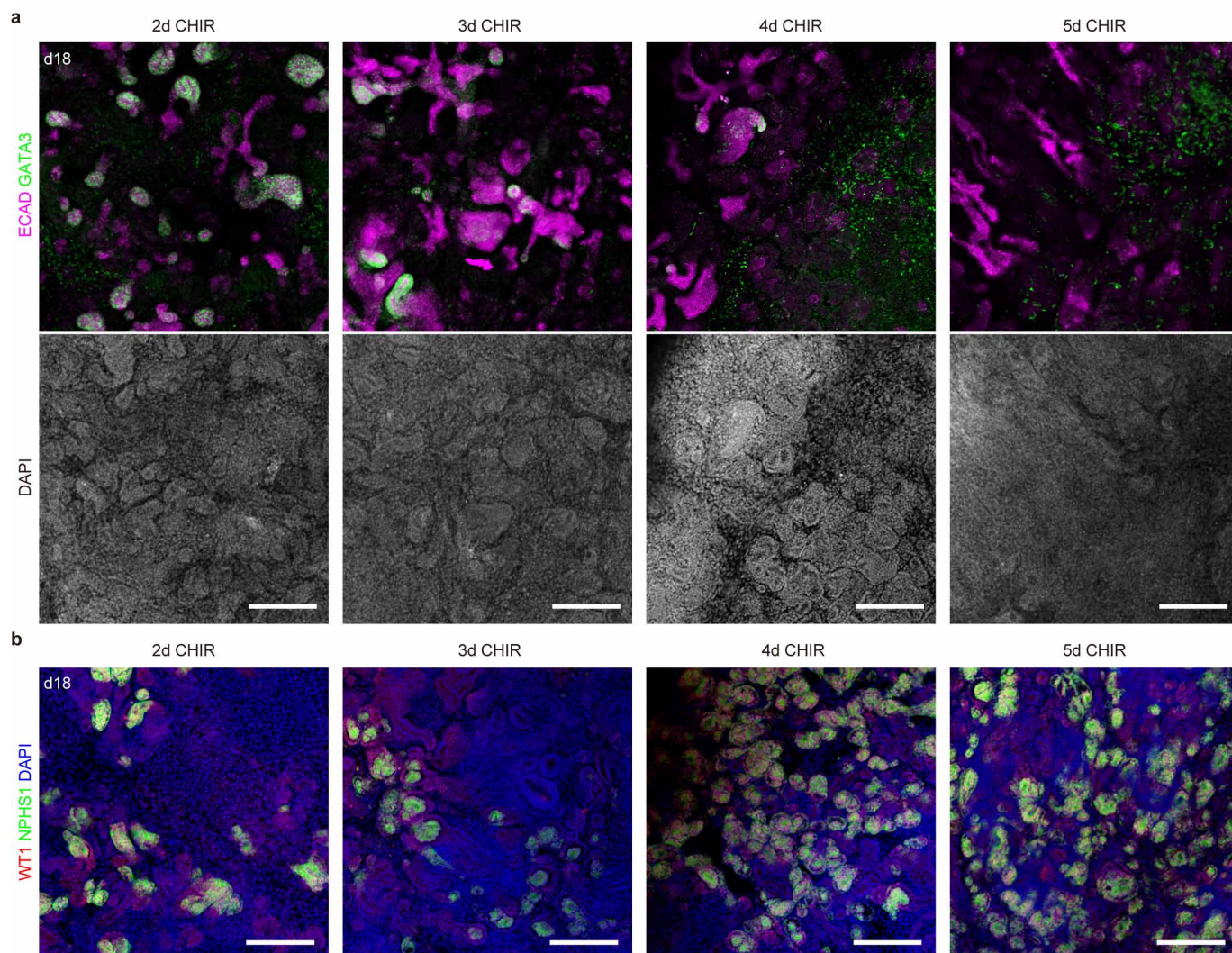
Extended Data Figure 4 | The timing of FGF9 exposure affects the ratio of collecting duct to nephron in the kidney organoid. a, **b**, Immunofluorescence of kidney organoids at day 18 after-aggregation after exposure to different timings of initial FGF9 exposure (2, 3, 4 and 5 days of CHIR99021 pre-FGF9), demonstrating the regulation of collecting duct/nephron ratio by varying this timing. GATA3$^+$ECAD$^+$ cells represent the collecting duct (**a**), whereas WT1$^+$NPHS1$^+$ cells mark podocytes of the glomerulus (**b**). Scale bars, 200 µm.

a

Nephrogenic mesenchyme

Ureteric tip

b

Early nephron

Proximal tubule

Podocytes

Distal tubule / Loop of Henle

c

Developing vasculature

Stroma / ECM
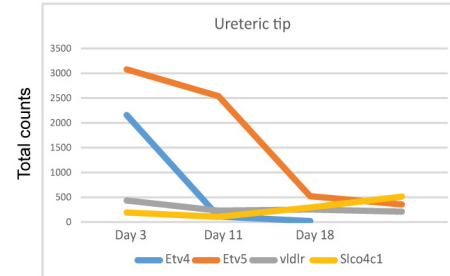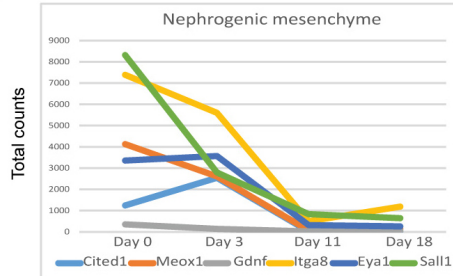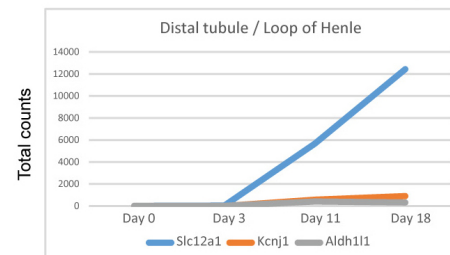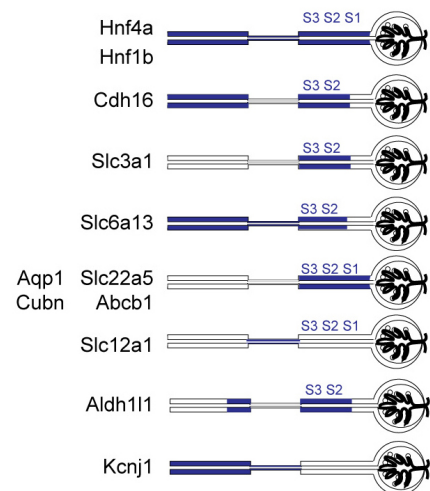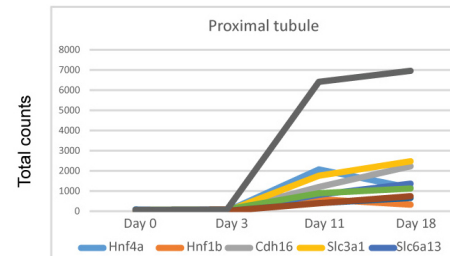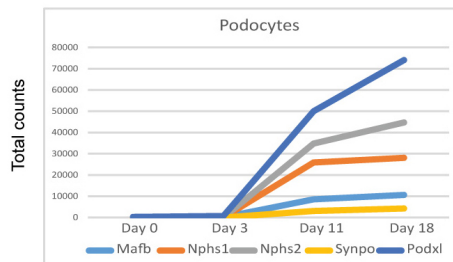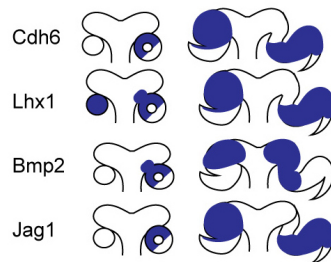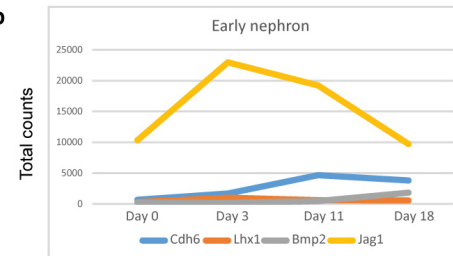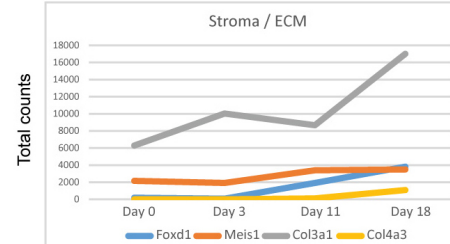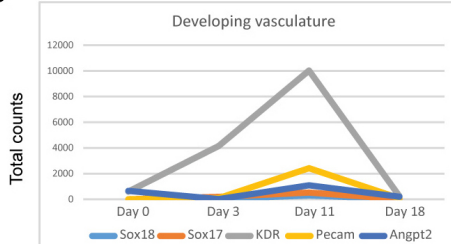
Extended Data Figure 5 | Changes of gene expression during development of the kidney organoid. a–c, Graphs showing expression changes of selected marker genes at 4 time points (day 0, 3, 11 and 18) of the kidney organoid culture. y axis represents the count of detection for each gene in an RNA sequencing analysis. Markers of the nephron progenitor (cap mesenchyme) and collecting duct progenitor (ureteric tip) were peaked by day 3 then dropped (**a**). Markers of early nephron increased by day 3, while those of mature nephron components (Proximal and distal tubule and Podocytes) started after day 3. Illustrations show expression regions (blue coloured) of each selected gene in the developing kidney (**b**). Markers of endothelial and renal interstitial cells were also increased by day 11 (**c**).

Height

0 1 2 3 4

DA2 pancreas_9
BP1 pancreas_9
CR1 spleen_22
CF2 spleen_22
CC2 spleen_18
CI1 spleen_16
CJ1 intestine_9
DF1 umbilical cord_9
CJ1 tongue_9
DF1 tongue_9
DJ1 muscle_9
CV1 eye_16
CJ1 umbilical cord_9
DF1 skin_9
DJ1 skin_9
DF1 eye_9
CJ1 eye_9
CH1 muscle_9
CJ1 spinal cord_9
DF1 spinal cord_9
CH1 adrenal_9
DJ1 brain_9
CC2 brain_18
CI1 brain_16
CR1 brain_22
DF1 brain_9
CV2 heart A_16
CC3 heart A_22
CF2 heart A_22
CV1 heart A_16
CI1 muscle_16
CF2 muscle_22
CV1 heart V_16
CJ1 heart V_9
CV2 heart V_16
DF1 heart V_9
CC3 heart V_22
CV1 muscle_16
CR1 muscle_22
CL3 gonad_9
DJ2 gonad_18
CH1 gonad_9
D0/D3
Day3.Pellet2_ORG
Day3.Pellet1_ORG
Day3.Pellet3_ORG
Day0.Pellet3_SC
Day0.Pellet4_SC
Day0.Pellet5_SC
CJ1 pancreas_9
Kidney 2T
CR1 kidney_22
CF2 kidney_22
CI1 kidney_16
CC2 kidney_18
D11/D18
Day11.Pellet3_ORG
Day11.Pellet1_ORG
Day11.Pellet2_ORG
Day18.Pellet1_ORG
Day18.Pellet2_ORG
Day18.Pellet5_ORG
Kidney 1T
DJ1 kidney_9
DF1 kidney_9
CJ1 amnion_9
DF1 amnion_9
CF2 amnion_22
CI1 amnion_16
CC2 amnion_18
CV1 amnion_16
CF2 stomach_22
DF1 stomach_9
CJ1 stomach_9
CC2 stomach_18
CV1 stomach_16
CR1 stomach_22
CR1 intestine_22
CC2 intestine_18
CI1 intestine_16
CP2 intestine_22
CF2 intestine_22
CC2 pancreas_18
CI1 pancreas_16
BO1 pancreas_22
CF2 pancreas_22
CI1 mother_16
CR1 mother_22
CV2 gonad_16
CC2 adrenal_18
CF2 adrenal_22
CI1 liver_16
CC2 liver_18
CR1 liver_22
DJ1 liver_9
DA2 liver_9
CF2 liver_22
CR1 skin_22
CV1 skin_16
CF2 skin_22
CI1 skin_16
CC2 skin_18
CF2 eye_22
CV2 eye_16
CP1 eye_22
CC2 tongue_18
CC2 tongue.2_18
CI1 tongue_16
CR1 tongue_22
BP1 lung_9
CF2 lung_22
DJ1 lung_9
CI1 lung_16
CC2 lung_18
CR1 lung_22
DF1 placenta_9
CJ1 placenta_9
CV1 placenta_16
CR1 placenta_22
CR1 amnion_22
CI1 chorion_16
CP1 chorion_22
CF2 placenta_22
CC2 placenta_18
CJ1 mother_9
CI1 placenta_16
CR1 chorion_22

Extended Data Figure 6 | Transcriptional similarity of the kidney organoid to human fetal organs. Dendrogram showing the hierarchical clustering of day 0, 3, 11 and 18 differentiation experiments and 21 human fetal organs from first and second trimester (Gene Expression Omnibus accession number GSE66302)[15]. Sample name is composed of individual ID followed by an organ name and gestation week. For instance, 'DJ1 kidney_9' represents a kidney at ninth week gestation from individual ID: DJ1. Day 0 and 3 kidney organoids cluster with gonad, in agreement with the common origin of both gonad and kidney from the intermediate mesoderm. Day 11 and 18 kidney organoids show strongest similarity to trimester 1 human kidney. The classifier genes used for this analysis are detailed in Supplementary Table 3.
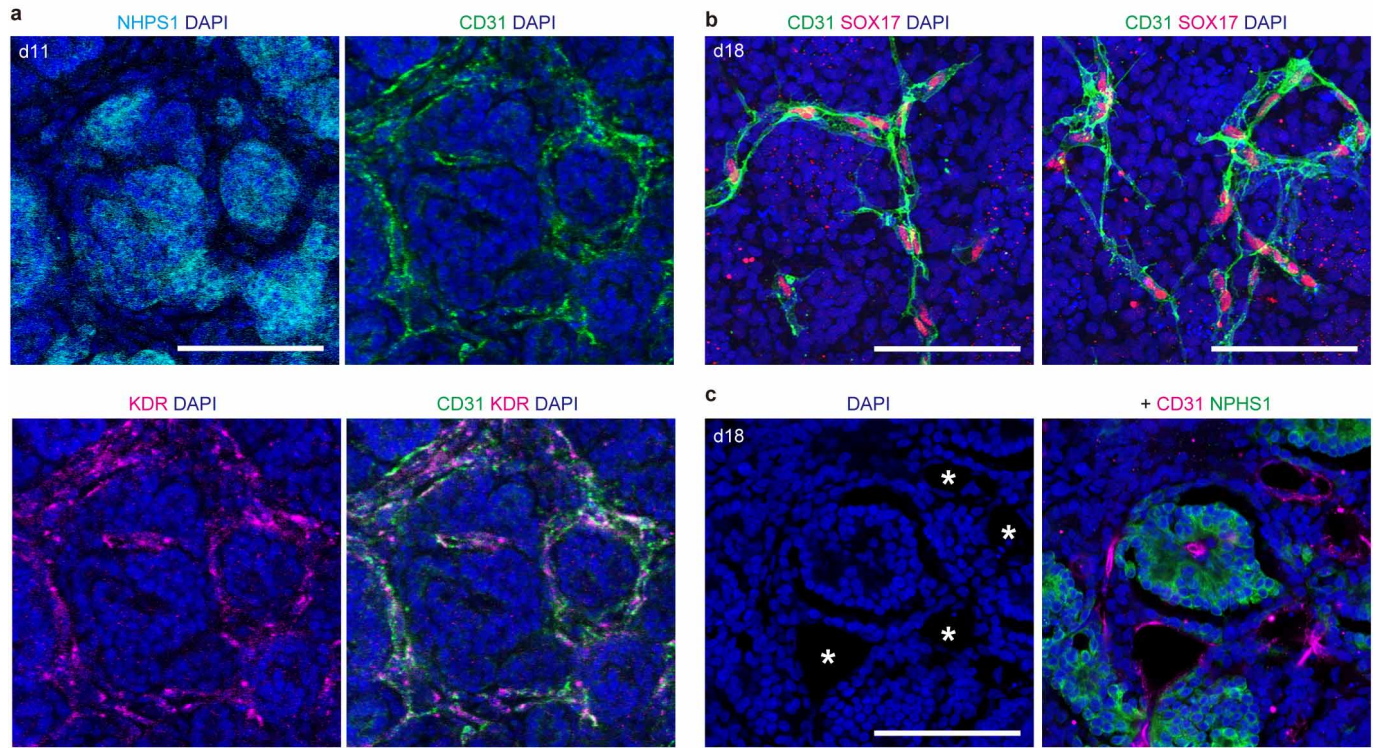
Extended Data Figure 7 | Evidence of endothelial cells in the kidney organoid. a, Immunofluorescence of day 11 kidney organoids showing the presence of CD31$^+$KDR$^+$ endothelial cells surrounding NPHS1$^+$ glomeruli. Scale bar, 100 μm. **b**, Two representative images demonstrating the expression of another endothelium marker SOX17 in CD31$^+$ endothelial cells. Scale bars, 100 μm. **c**, Immunofluorescence of day 18 kidney organoids displaying endothelia with lumen formation, as indicated by asterisks. This image also shows the endothelial invasion into a glomerulus. Scale bar, 100 μm.
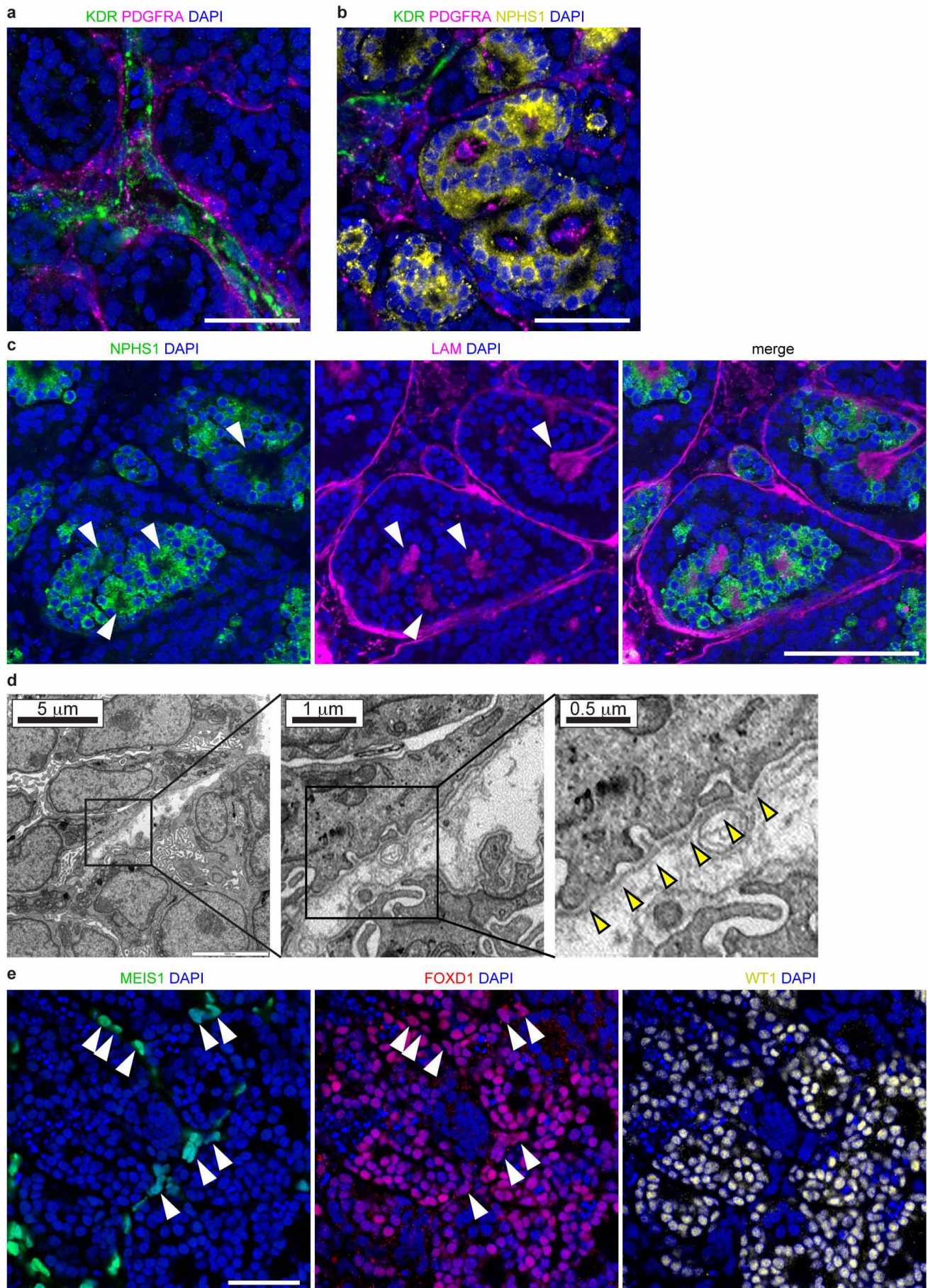
Extended Data Figure 8 | Characterization of non-epithelial structures in the kidney organoid. All images were taken from day 18 kidney organoids. **a**, PDGFRA$^+$ pericytic cells attaching on KDR$^+$ vessels. Scale bar, 50 μm. **b**, Some glomeruli contained PDGFRA$^+$ cells likely to represent early mesangial cells[19]. Scale bar, 50 μm. **c**, Laminin staining (LAM) demonstrates the presence of basement membrane in glomerulus structures (white arrowheads). Scale bar, 100 μm. **d**, TEM images of avascular glomeruli showing early podocytes surrounding a basement membrane (yellow arrowheads) and exhibiting foot processes on the basement membrane. **e**, Immunofluorescence showing FOXD1 expression in podocytes (WT1$^+$FOXD1$^+$)[18] and a subpopulation of MEIS1$^+$ interstitium (white arrowheads). This is suggestive of the presence of both cortical stroma (FOXD1$^+$MEIS1$^+$) and medullary stroma (FOXD1$^-$MEIS1$^+$). Scale bar, 100 μm.
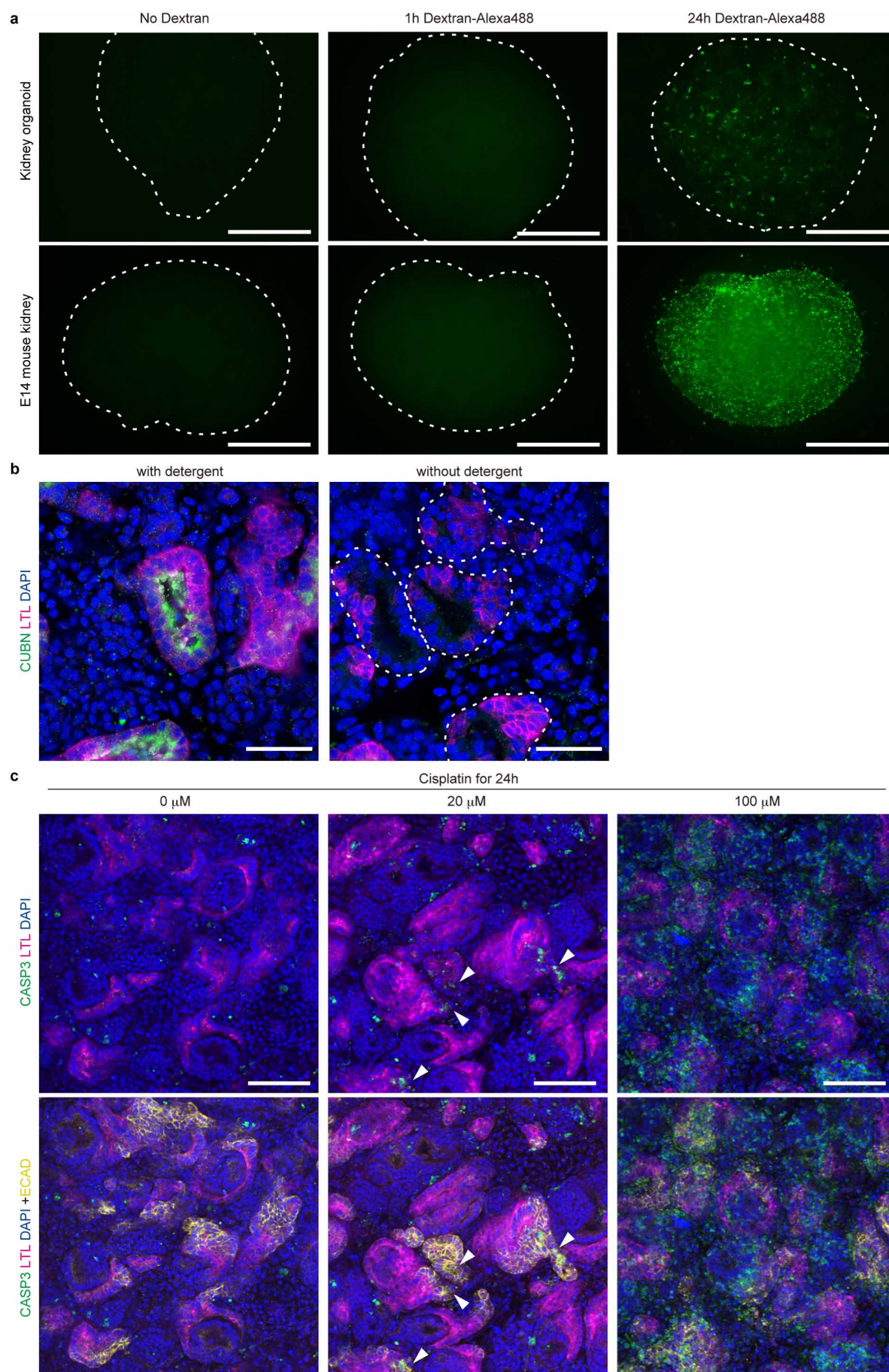
Extended Data Figure 9 | Functional assay of proximal tubule maturation within kidney organoids. a, Fluorescent microscopy showing the dextran uptake in both the kidney organoids and E14 mouse embryonic kidneys organ culture after 24 h presence of dextran–Alexa488 ($10\,\mu g\,ml^{-1}$) in the culture medium (24 h dextran–Alexa488). 1 h incubation was insufficient for either organoids or mouse kidney explants to uptake dextran from the culture media (1 h dextran–Alexa488). No background signals were detected in a control without dextran (no dextran). Dashed line circles the organoids and kidneys. Scale bars, 1 mm. **b**, Endocytosis mediator cubilin (CUBN) was present on apical surface of the proximal tubules in kidney organoids (left panel). The same staining without detergent during the process showed the complete absence of CUBN staining on apical surface (right panel), demonstrating that the tubules within the organoids are intact. This explains the requirement for a 24 h incubation with dextran before evidence of apical uptake. Dashed line circles LTL^{+} proximal tubules. Scale bars, 50 μm. **c**, Low power immunofluorescence microscopy of day 18 kidney organoids after being treated by cisplatin for 24 h. No apoptosis was observed in proximal tubules in the absence of cisplatin (0 μM, left panel). LTL^{+}ECAD^{+} proximal tubular cell-specific apoptosis was observed only in response to either 5 μM (not shown) or 20 μM cisplatin (arrowheads in middle panel). Global cell death was observed after culture in 100 μM cisplatin (right panel). Scale bars, 100 μm.

LETTER

A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing

Richard O. Prum[1,2]*, Jacob S. Berv[3]*, Alex Dornburg[1,2,4], Daniel J. Field[2,5], Jeffrey P. Townsend[1,6], Emily Moriarty Lemmon[7] & Alan R. Lemmon[8]

Although reconstruction of the phylogeny of living birds has progressed tremendously in the last decade, the evolutionary history of Neoaves—a clade that encompasses nearly all living bird species—remains the greatest unresolved challenge in dinosaur systematics. Here we investigate avian phylogeny with an unprecedented scale of data: >390,000 bases of genomic sequence data from each of 198 species of living birds, representing all major avian lineages, and two crocodilian outgroups. Sequence data were collected using anchored hybrid enrichment, yielding 259 nuclear loci with an average length of 1,523 bases for a total data set of over 7.8×10^7 bases. Bayesian and maximum likelihood analyses yielded highly supported and nearly identical phylogenetic trees for all major avian lineages. Five major clades form successive sister groups to the rest of Neoaves: (1) a clade including nightjars, other caprimulgiforms, swifts, and hummingbirds; (2) a clade uniting cuckoos, bustards, and turacos with pigeons, mesites, and sandgrouse; (3) cranes and their relatives; (4) a comprehensive waterbird clade, including all diving, wading, and shorebirds; and (5) a comprehensive landbird clade with the enigmatic hoatzin (*Opisthocomus hoazin*) as the sister group to the rest. Neither of the two main, recently proposed Neoavian clades—Columbea and Passerea[1]—were supported as monophyletic. The results of our divergence time analyses are congruent with the palaeontological record, supporting a major radiation of crown birds in the wake of the Cretaceous–Palaeogene (K–Pg) mass extinction.

Birds (Aves) are the most diverse lineage of extant tetrapod vertebrates. They comprise over 10,000 living species[2], and exhibit an extraordinary diversity in morphology, ecology, and behaviour[3]. Substantial progress has been made in resolving the phylogenetic history of birds. Phylogenetic analyses of both molecular and morphological data support the monophyletic Palaeognathae (the tinamous and flightless ratites) and Galloanserae (gamebirds and waterfowl) as successive, monophyletic sister groups to the Neoaves—a diverse clade including all other living birds[4]. Resolving neoavian phylogeny has proven to be a difficult challenge because this radiation was very rapid and deep in time, resulting in very short internodes[4].

In the last decade, phylogenetic analyses of large, multilocus data sets have resulted in the proposal of numerous, novel neoavian relationships. For example, a clade consisting of diving and wading birds has been consistently recovered, as well as a large landbird clade in which falcons and parrots are successive sister groups to the perching birds[4–8]. Recently, phylogenetic analyses of 48 whole avian genomes resulted in the proposal of a novel phylogenetic resolution of the initial branching sequence within Neoaves[1]. Although this genomic study provided much needed corroboration of many neoavian clades, the limited taxon sampling precluded further insights into the evolutionary history of birds.

It has long been recognized that phylogenetic confidence depends not only on the number of characters analysed and their rate of evolution, but also on the number and relationships of the taxa sampled relative to the nodes of interest[9–11]. Theory predicts that sampling a single taxon that diverges close to a node of interest will have a far greater effect on phylogenetic resolution than will adding more characters[11]. Despite using an alignment of >40 million base pairs, sparse sampling of 48 species in the recent avian genomic analysis may not have been sufficient to confidently resolve the deep divergences among major lineages of Neoaves. Thus, expanded taxon sampling is required to test the monophyly of neoavian clades, and to further resolve the phylogenetic relationships within Neoaves.

Here, we present a phylogenetic analysis of 198 bird species and 2 crocodilians (Supplementary Table 1) based on loci captured using anchored enrichment[12]. Our sample includes species of 122 avian families in all 40 extant avian orders[2], with denser representation of non-oscine birds (108 families) than of oscine songbirds (14 families). Effort was made to include taxa that would break up long phylogenetic branches, and provide the highest likelihood of resolving short internodes at the base of Neoaves[11]. We also sampled multiple species within groups whose monophyly or phylogenetic interrelationships have been controversial—that is, tinamous, nightjars, hummingbirds, turacos, cuckoos, pigeons, sandgrouse, mesites, rails, storm petrels, petrels, storks, herons, hawks, hornbills, mousebirds, trogons, kingfishers, barbets, seriemas, falcons, parrots, and suboscine passerines.

We targeted 394 loci centred on conserved anchor regions of the genome that are flanked by more variable regions[12]. We performed all phylogenetic analyses on a data set of 259 genes with the highest quality assemblies. The average locus was 1,524 bases in length (361–2,316 base pairs (bp)), and the total percentage of missing data was 1.84%. The concatenated alignment contained 394,684 sites. To minimize overall model complexity while accurately accounting for substitution processes, we performed a partition model sensitivity analysis with PartitionFinder[13,14], and compared a complex partition model (one partition per locus) to a heuristically optimized (rclust) partition model. Phylogenetic informativeness (PI) approaches[15,16] provided strong evidence that the phylogenetic utility of our data set was high, with low declines in PI profiles for individual loci, data set partitions, and the concatenated matrix (Supplementary Fig. 4). We estimated concatenated trees in ExaBayes[17] and RAxML[18] using a 75 partition model. Coalescent species trees were estimated with the gene tree summation methods in STAR[19], NJst[20], and ASTRAL[21] from gene trees estimated with RAxML (see Methods.)

Our concatenated Bayesian analyses resulted in a completely resolved, well supported phylogeny. All clades had a posterior probability (PP) of 1, except for a single clade including shoebill (*Balaeniceps*) and pelican (PP = 0.54) (Fig. 1). The concatenated

[1]Department of Ecology & Evolutionary Biology, Yale University, New Haven, Connecticut 06520, USA. [2]Peabody Museum of Natural History, Yale University, New Haven, Connecticut 06520, USA. [3]Department of Ecology and Evolutionary Biology, Fuller Evolutionary Biology Program, Cornell University, and Cornell Laboratory of Ornithology, Ithaca, New York 14853, USA. [4]North Carolina Museum of Natural Sciences, Raleigh, North Carolina 27601, USA. [5]Department of Geology & Geophysics, Yale University, New Haven, Connecticut 06520, USA. [6]Department of Biostatistics, and Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA. [7]Department of Biological Science, Florida State University, Tallahassee, Florida 32306, USA. [8]Department of Scientific Computing, Florida State University, Tallahassee, Florida 32306, USA.
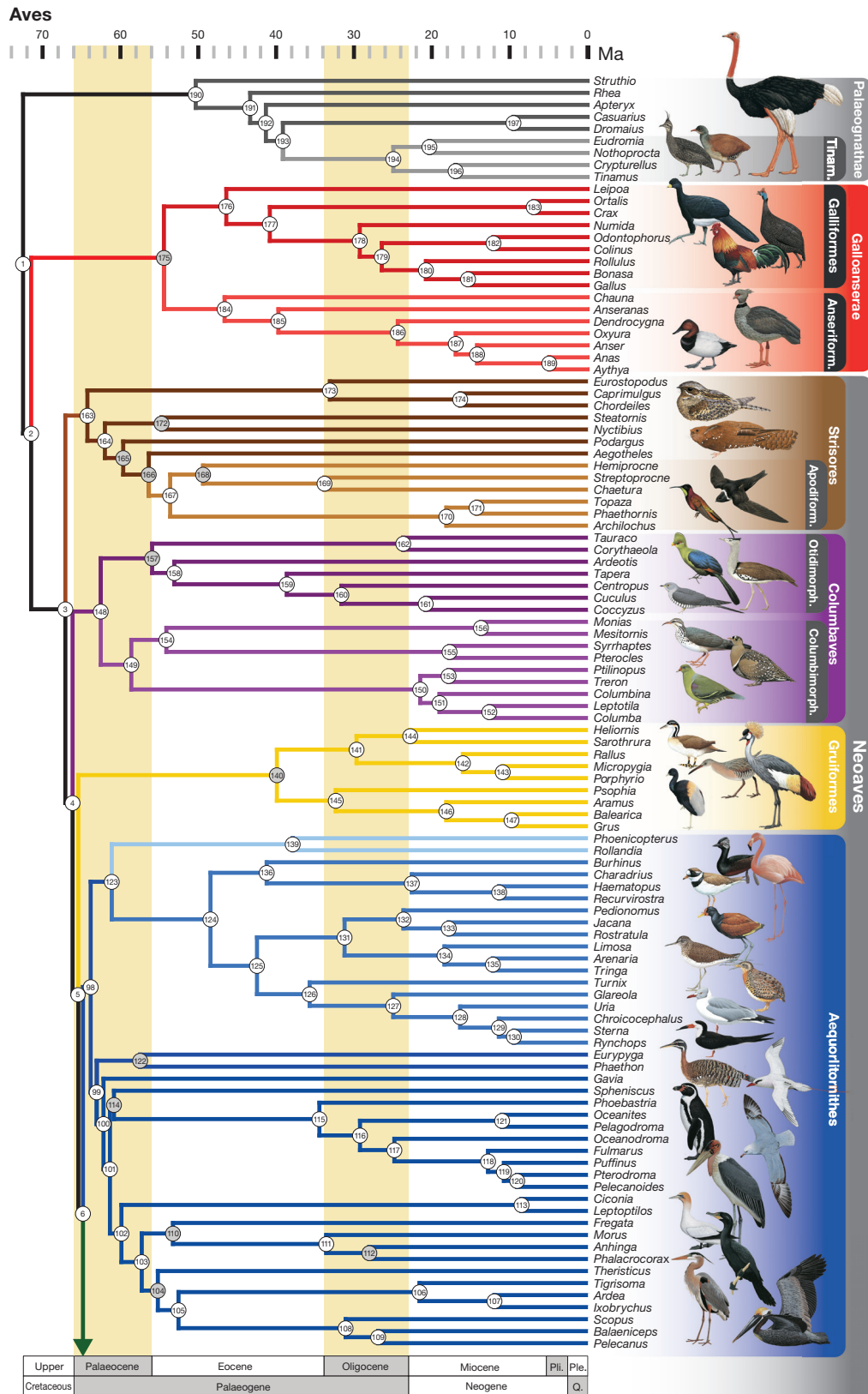*These authors contributed equally to this work.

Figure 1 | Phylogeny of birds. Time-calibrated phylogeny of 198 species of birds inferred from a concatenated, Bayesian analysis of 259 anchored phylogenomic loci using ExaBayes[17]. Figure continues on the opposite page from green arrow at the bottom of this panel. Complete taxon data in Supplementary Table 1. Higher taxon names appear at right. All clades are supported with posterior probability (PP) of 1.0, except for the *Balaeniceps–Pelecanus* clade (PP = 0.54; clade 109). The five major, successive, neoavian sister clades are: Strisores (brown), Columbaves (purple), Gruiformes (yellow), Aequorlitornithes (blue), and Inopinaves (green). Background colours mark geological periods. Ma, million years ago; Ple, Pleistocene; Pli, Pliocene; Q., Quaternary. Clade numbers refer to the plot of estimated divergence dates (Supplementary Fig. 7). Fossil age-calibrated nodes are shown in grey. Illustrations of representative bird species[30] are depicted by their lineages. See Supplementary Information for details and further discussion.
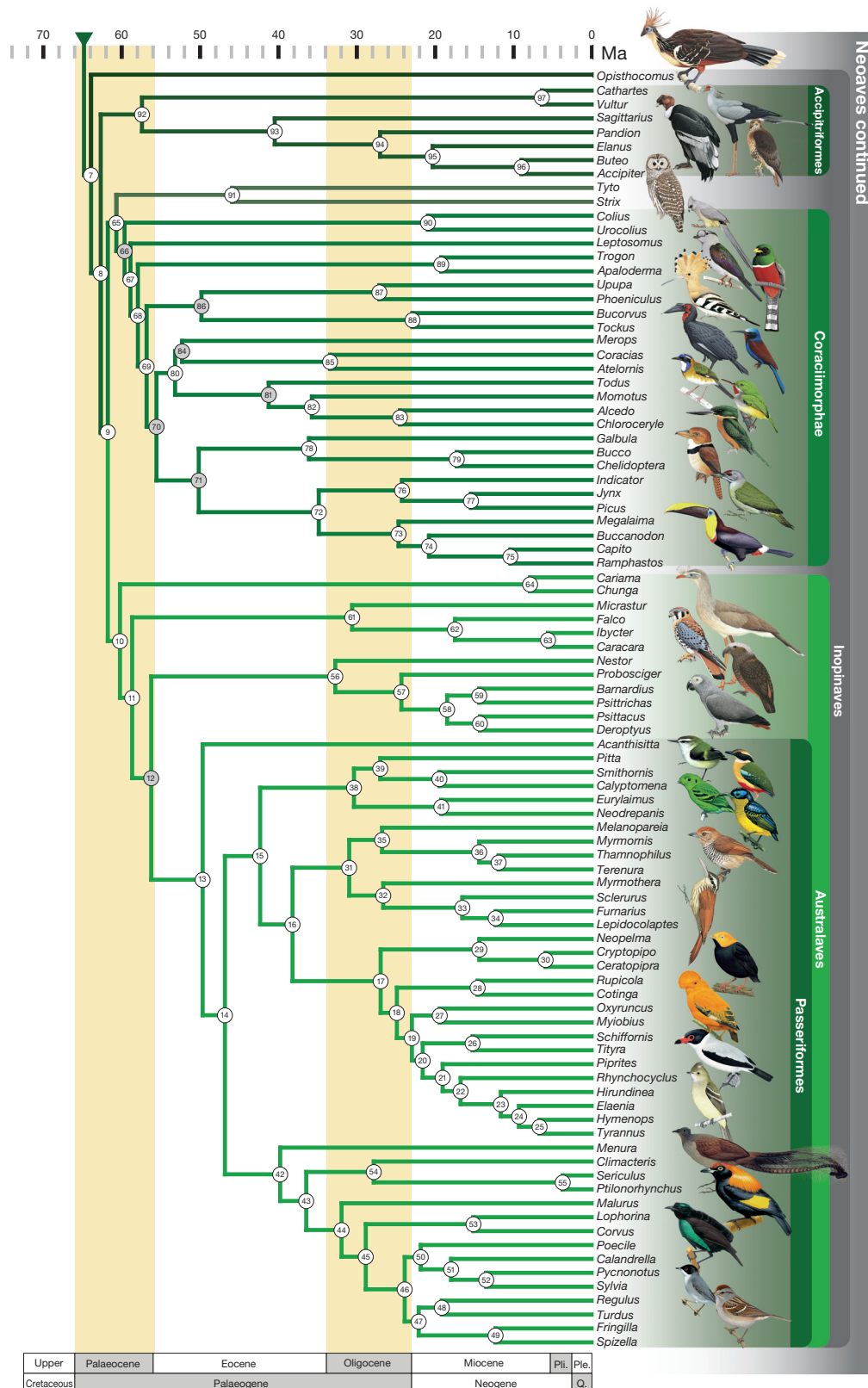
Figure 1 | Continued.

maximum likelihood analysis recovered a single topology that was identical to the Bayesian tree except for three clades, all of which are far from the base of Neoaves: the relationships among pigeons; among skimmers, gulls, and terns; and among pelicans, shoebill, and waders (Supplementary Fig. 1). Almost all clades in the maximum likelihood

tree were maximally supported with bootstrap scores (BS) of 1.00, but nine clades within Neoaves (including four of the most inclusive neoavian clades) received support <0.70 (Supplementary Fig. 1). Coalescent species tree analyses produced substantially different hypotheses for neoavian relationships (Supplementary Fig. 3), but

most of the discordant clades received conspicuously lower bootstrap support values ($0.07 < BS < 0.30$). Quantifying the phylogenetic informativeness of individual loci[15,16] revealed that these low support values were not due to homoplasy driven by saturation of nucleotide states, but rather by the low power of individual loci to resolve the entire range of internode lengths across the depth of the tree (Supplementary Figs 4 and 5; see Methods). This result was not unexpected. The low phylogenetic information content of individual genes at deep timescales has been demonstrated to impede phylogenetic resolution in a coalescent species tree framework[22,23]. Furthermore, when clades with <0.75 bootstrap support values in the species trees are collapsed, the resulting topology is exactly congruent with the concatenated Bayesian tree (except for the relationships of tinamous among palaeognaths; Supplementary Fig. 3). Although coalescent species trees account for incomplete lineage sorting, simulations show that species tree methods based on gene tree summation may not provide significantly better performance over concatenation methods[22].

Our phylogeny identifies many new clades, and supports many phylogenetic relationships proposed in previous studies (see detailed phylogenetic discussion in the Supplementary Information). Congruent with all recent studies, the phylogeny places palaeognaths as the sister group to the rest of birds, and the flying tinamous (Tinamidae) within the flightless ratites. This tree, however, places tinamous as the sister group to cassowary and emu alone (Fig. 1, grey). The phylogeny of Galloanserae is exactly congruent with previous studies[4] (Fig. 1, red).

Within the monophyletic Neoaves, we recover five major clades, each of which is the successive sister group to the remaining clades in the series (Fig. 1). The Strisores includes the nightjars and their nocturnal relatives with the diurnal swifts and hummingbirds (Fig. 1, brown). Four nocturnal lineages—nightjars, a neotropical oilbird-potoo clade, frogmouths, and owlet-nightjars—form successive sister groups to the diurnal swift and hummingbird clade.

The Columbaves is a novel clade that consists of two monophyletic groups recently identified by Jarvis et al.[1] (Fig. 1, purple). A clade consisting of turacos, bustards, and cuckoos (Otidimorphae) is sister to a clade consisting of pigeons as the sister group to sandgrouse and mesites (Columbimorphae). The third neoavian clade consists of a well recognized monophyletic group of core gruiform birds (Gruiformes; Fig. 1, yellow), with interrelationships that are consistent with previous phylogenies[4].

The Aequorlitornithes is a novel, comprehensive clade of waterbirds, including all shorebirds, diving birds, and wading birds (Fig. 1, blue). Within this group, the flamingos and grebes[1,4–6] are the sister group to shorebirds, and the sunbittern and tropicbirds[1,4,6] are the sister group to the wading and diving birds (Fig. 1, blue). Other interrelationships within these groups are extensively congruent with the results in ref. 4 and the work of others (see Supplementary Information).

The fifth major neoavian clade, which we name Inopinaves, is a very diverse landbird clade with the same composition as previously recognized (Telluraves)[1,4–6], but with the enigmatic, neotropical hoatzin (Opisthocomus hoazin) as the sister group to all other landbirds (Fig. 1, green). The phylogeny of the landbirds shares many points of congruence with earlier hypotheses, including the relationships of seriemas, falcons, parrots, and perching birds[1,4–6], and the interrelationships among oscine songbirds[24]. However, we find that hawks (Accipitriformes) are the sister group to a new clade including the rest of the landbirds, to be called Eutelluraves (see Supplementary Information).

Our divergence time analyses employed 19 phylogenetically and geologically well-constrained fossil calibrations (following recently proposed best practices[25]), documenting many deep divergences within the avian crown group (Fig. 1, grey nodes; see Supplementary Information). Our analysis supports an extremely rapid radiation of the avian crown group in the wake of the K–Pg mass extinction event (Fig. 1, Supplementary Figs 6 and 7). Although the post-K–Pg radiation hypothesis has long been strongly supported by the avian fossil

record[26,27], it has so far received little support from molecular divergence time analyses[4,28]. The tempo and mode of the extant avian radiation remains contentious. For example, an alternative calibration analysis including the fossil Vegavis did not support significantly different dates of divergence outside of the Galloanserae (see Supplementary Information and Supplementary Figs 10–12). Confident determination of the age of crown Aves will have to await discoveries of Mesozoic stem neognaths and palaeognaths, and detailed assessments of the influence of soft maximum bound parameterization on the age of the deepest avian divergences.

Our results indicate that the recent genome phylogeny[1] may contain some erroneous relationships induced by long branch attraction from sparse taxon sampling. Maximum likelihood analysis of our sequence data pruned down to a phylogenetically equivalent subsample of 48 species produces relationships along the neoavian 'backbone' (Supplementary Fig. 8) that are entirely discordant with the phylogeny based on our full data set (Fig. 1). This reduced taxon analysis recovers some of the specific features of the recent genome phylogeny by Jarvis et al.[1] (Supplementary Fig. 8): for example, the placement of the pigeons, mesites, and sandgrouse (a subclade of Columbea[1]) outside of the rest of the Neoaves. Differences in tree topology when taxa are excluded are to be expected if early internodes in Neoaves are very short. Adding taxa that have diverged near nodes of interest has been theoretically demonstrated to constrain the possible historical substitution patterns, and increase the accuracy of phylogenetic inference[11]. By increasing our taxon sampling to include all major avian lineages, we have minimized the possibility that additional taxon sampling alone will alter the relationships in our tree.

Jarvis et al.[1] also identified a well supported clade consisting of the hoatzin (Opisthocomus) as the sister group to a crane (Grus) and a plover (Charadrius) (total evidence nucleotide tree, BS = 0.91, 0.96, respectively). However, Grus and Charadrius were the only species sampled from two very diverse neoavian orders: Gruiformes, 185 species; and Charadriiformes, 385 species[2]. Our results indicate that Opisthocomus is the most ancient bird lineage (~ 64 million years) consisting of only a single, extant species. Thus, the three taxa placed in this assemblage by Jarvis et al.[1] comprise three of the most ancient, and under-sampled lineages within all birds, indicating the strong possibility of long branch attraction artefacts. By contrast, these same groups are represented by 26 species in our analysis, and they do not form an exclusive clade (Fig. 1).

In addition to providing a new backbone for comprehensive avian supertrees and comparative evolutionary analyses[28], this new avian phylogeny supports many interesting hypotheses about avian evolution. This phylogeny upholds the hypothesis that the ancestor of the diurnal swifts and hummingbirds evolved from a clade that had been predominantly nocturnal for ~ 10 million years. Although hummingbirds have acute near-ultraviolet vision[29], the effect of extended ancestral nocturnality on the evolution of the visual system in this group of birds is unknown. Our findings also support the emerging pattern that landbirds evolved from a raptorial grade[1]. The sister group relationships of hawks to the rest of the landbirds, of owls to the diverse coraciimorph clade, and of seriemas and falcons to the parrots and passerines indicate the persistence of a raptorial ecology among ancestral landbirds. Lastly, the identification of a new, broadly comprehensive waterbird–shorebird clade indicates a striking, and previously unappreciated, level of evolutionary constraint on the ecological diversification of birds that will be exciting to investigate in the future.

1. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science **346,** 1320–1331 (2014).

2. Gill, F. & Donsker, D. IOC World Bird List (v5.1) http://dx.doi.org/10.14344/IOC.ML.5.1 (2015).
3. Gill, F. B. *Ornithology* 2nd edn (W. H. Freeman and Co., 1995).
4. Hackett, S. J. *et al.* A phylogenomic study of birds reveals their evolutionary history. *Science* **320,** 1763–1768 (2008).
5. Ericson, P. G. P. *et al.* Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol. Lett.* **2,** 543–547 (2006).
6. McCormack, J. E. *et al.* A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE* **8,** e54848 (2013).
7. Mayr, G. *Paleogene Fossil Birds* (Springer, 2009).
8. Mayr, G. Metaves, Mirandornithes, Strisores and other novelties — a critical review of the higher-level phylogeny of neornithine birds. *J. Zoological Syst. Evol. Res.* **49,** 58–76 (2011).
9. Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47,** 9–17 (1998).
10. Heath, T. A., Hedtke, S. M. & Hillis, D. M. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* **46,** 239–257 (2008).
11. Townsend, J. P. & Lopez-Giraldez, F. Optimal selection of gene and ingroup taxon sampling for resolving phylogenetic relationships. *Syst. Biol.* **59,** 446–457 (2010).
12. Lemmon, A. R., Emme, S. A. & Lemmon, E. M. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* **61,** 727–744 (2012).
13. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29,** 1695–1701 (2012).
14. Berv, J. S. & Prum, R. O. A comprehensive multilocus phylogeny of the neotropical cotingas (Cotingidae, Aves) with a comparative evolutionary analysis of breeding system and plumage dimorphism and a revised phylogenetic classification. *Mol. Phylogenet. Evol.* **81,** 120–136 (2014).
15. Townsend, J. P. Profiling phylogenetic informativeness. *Syst. Biol.* **56,** 222–231 (2007).
16. Townsend, J. P., Su, Z. & Tekle, Y. I. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. *Syst. Biol.* **61,** 835–849 (2012).
17. Aberer, A. J., Kobert, K. & Stamatakis, A. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* **31,** 2553–2556 (2014).
18. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
19. Liu, L., Yu, L., Pearl, D. K. & Edwards, S. V. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* **58,** 468–477 (2009).
20. Liu, L. & Yu, L. Estimating species trees from unrooted gene trees. *Syst. Biol.* **60,** 661–667 (2011).
21. Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30,** i541–i548 (2014).
22. Tonini, J., Moore, A., Stern, D., Shcheglovitova, M. & Ortí, G. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLOS Currents Tree of Life* **1,** http://dx.doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be (2015).
23. Mirarab, S., Bayzid, M. S. & Warnow, T. Evaluating summary methods for multi-locus species tree estimation in the presence of incomplete lineage sorting. *Syst. Biol.* http://dx.doi.org/10.1093/sysbio/syu063 (2014).
24. Barker, F. K., Cibois, A., Schikler, P., Felsenstein, J. & Cracraft, J. Phylogeny and diversification of the largest avian radiation. *Proc. Natl Acad. Sci. USA* **101,** 11040–11045 (2004).
25. Parham, J. F. *et al.* Best practices for justifying fossil calibrations. *Syst. Biol.* **61,** 346–359 (2012).
26. Longrich, N. R., Tokaryk, T. & Field, D. J. Mass extinction of birds at the Cretaceous–Paleogene (K–Pg) boundary. *Proc. Natl Acad. Sci. USA* **108,** 15253–15257 (2011).
27. Feduccia, A. *The Origin and Evolution of Birds* 2nd edn (Yale Univ. Press, 1999).
28. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491,** 444–448 (2012).
29. Goldsmith, T. H. Hummingbirds see near ultraviolet light. *Science* **207,** 786–788 (1980).
30. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E. *Handbook of the Birds of the World Alive* (Lynx Edicions, 2015).

METHODS

Locus selection and probe design. Anchor loci described in ref. 12 were extended such that each contained approximately 1,350 bp. In some cases neighbouring loci were joined to form a single locus. Also, loci that performed poorly in ref. 12 were removed from the locus set. This process produced 394 loci (referred to as the version 2 vertebrate loci). Genome coordinates corresponding to these regions in the *Gallus gallus* genome (galGal3, UCSC genome browser) were identified and sequences corresponding to this region were extracted (coordinates are available in the Zenodo archive (http://dx.doi.org/10.5281/zenodo.28343)). In order to improve the capture efficiency for passerines, we also obtained homologous sequences for *Taeniopygia guttata*. After aligning the *Gallus* and *Taeniopygia* sequences using MAFFT[31], alignments were trimmed to produce the final probe region alignments (alignments available in the Zenodo archive), and probes were tiled at approximately $1.5\times$ tiling density (probe specification will be made available upon publication).

Data collection. Data were collected following the general methods of ref. 12 through the Center for Anchored Phylogenomics at Florida State University (http://www.anchoredphylogeny.com). Briefly, each genomic DNA sample was sonicated to a fragment size of ~150–350 bp using a Covaris E220 focused-ultrasonicator with Covaris microTUBES. Subsequently, library preparation and indexing were performed on a Beckman-Coulter Biomek FXp liquid-handling robot following a protocol modified from ref. 32. One important modification is a size-selection step after blunt-end repair using SPRIselect beads (Beckman-Coulter; $0.9\times$ ratio of bead to sample volume). Indexed samples were then pooled at equal quantities (typically 12–16 samples per pool), and enrichments were performed on each multi-sample pool using an Agilent Custom SureSelect kit (Agilent Technologies), designed as specified above. After enrichment, the 12 enrichment pools were pooled in groups of three in equal quantities for sequencing on four PE150 Illumina HiSeq2000 lanes (three enrichment pools per lane). Sequencing was performed in the Translational Science Laboratory in the College of Medicine at Florida State University.

Data processing. *Paired-read merging* (Merge.java). Typically, between 50% and 75% of sequenced library fragments had an insert size between 150 bp and 300 bp. As 150 bp paired-end sequencing was performed, this means that the majority of the paired reads overlap and thus should be merged before assembly. The overlapping reads were identified and merged following the methods of ref. 33. In short, for each degree of overlap for each read we computed the probability of obtaining the observed number of matches by chance, and selected degree of overlap that produced the lowest probability, with a *P* value less than 10^{-10} required to merge reads. When reads are merged, mismatches are reconciled using base-specific quality scores, which were combined to form the new quality scores for the merged read (see ref. 33 for details). Reads failing to meet the probability criterion were kept separate but still used in the assembly. The merging process produces three files: one containing merged reads and two containing the unmerged reads.

Assembly (Assembler.java). The reads were assembled into contigs using an assembler that makes use of both a divergent reference assembly approach to map reads to the probe regions and a *de novo* assembly approach to extend the assembly into the flanks. The reference assembler uses a library of spaced 20-mers derived from the conserved sites of the alignments used during probe design. A preliminary match was called if at least 17 of 20 matches exist between a spaced kmer and the corresponding positions in a read. Reads obtaining a preliminary match were then compared to an appropriate reference sequence used for probe design to determine the maximum number of matches out of 100 consecutive bases (all possible gap-free alignments between the read and the reference ware considered). The read was considered mapped to the given locus if at least 55 matches were found. Once a read is mapped, an approximate alignment position was estimated using the position of the spaced 20-mer, and all 60-mers existing in the read were stored in a hash table used by the *de novo* assembler. The *de novo* assembler identifies exact matches between a read and one of the 60-mers found in the hash table. Simultaneously using the two levels of assembly described above, the three read files were traversed repeatedly until an entire pass through the reads produced no additional mapped reads.

For each locus, mapped reads were then clustered into clusters using 60-mer pairs observed in the reads mapped to that locus. In short, a list of all 60-mers found in the mapped reads was compiled, and the 60-mers were clustered if found together in at least two reads. The 60-mer clusters were then used to separate the reads into clusters for contig estimation. Relative alignment positions of reads within each cluster were then refined in order to increase the agreement across the reads. Up to one gap was also inserted per read if needed to improve the alignment. Note that given sufficient coverage and an absence of contamination, each single-copy locus should produce a single assembly cluster. Low coverage (leading to a break in the assembly), contamination, and gene duplication, can all

lead to an increased number of assembly clusters. A whole-genome duplication, for example, would increase the number of clusters to two per locus.

Consensus bases were called from assembly clusters as follows. For each site an unambiguous base was called if the bases present were identical or if the polymorphism of that site could be explained as sequencing error, assuming a binomial probability model with the probability of error equal to 0.1 and alpha equal to 0.05. If the polymorphism could not be explained as sequencing error, the ambiguous base was called that corresponded to all of the observed bases at that site (for example, 'R' was used if 'A' and 'G' were observed). Called bases were soft-masked (made lowercase) for sites with coverage lower than five. A summary of the assembly results is presented in a spreadsheet in the electronic data archive (http://dx.doi.org/10.5281/zenodo.28343; Prum_AssemblySummary_Summary.xlsx).

Contamination filtering (IdentifyGoodSeqsViaReadsMapped.r, GatherALL ConSeqsWithOKCoverage.java). In order to filter out possible low-level contaminants, consensus sequences derived from very low coverage assembly clusters (<10 reads) were removed from further analysis. After filtering, consensus sequences were grouped by locus (across individuals) in order to produce sets of homologues.

Orthology (GetPairwiseDistanceMeasures.java, plotMDS5.r). Orthology was then determined for each locus as follows. First, a pairwise distance measure was computed for pairs of homologues. To compute the pairwise distance between two sequences, we computed the percent of 20-mers observed in the two sequences that were found in both sequences. Note that the list of 20-mers was constructed from consecutive 20-mers as well as spaced 20-mers (every third base), in order to allow increased levels of sequence divergence. Using the distance matrix, we clustered the sequences using a neighbour-joining algorithm, but allowing at most one sequence per species to be in a given cluster. Clusters containing fewer than 50% of the species were removed from downstream processing.

Alignment (MAFFT). Sequences in each orthologous set were aligned using MAFFT v7.023b[31] with "−genafpair" and "−maxiterate 1000" flags.

Alignment Trimming (TrimAndMaskRawAlignments3). The alignment for each locus was then trimmed/masked using the following procedure. First, each alignment site was identified as 'good' if the most common character observed was present in >40% of the sequences. Second, 20 bp regions of each sequence that contained <10 good sites were masked. Third, sites with fewer than 12 unmasked bases were removed from the alignment. Lastly, entire loci were removed if both outgroups or more than 40 taxa were missing. This filter yielded 259 trimmed loci containing fewer than 2.5% missing characters overall.

Model selection and phylogenetic inference. To minimize the overall model complexity while accurately accounting for substitution processes, we performed a partition-model sensitivity analysis with the development version of PartitionFinder v2.0 (ref. 13), sensu[14], and compared a complex partition-model (one partition per gene) to a heuristically optimized (relaxed clustering with the RAxML option for accelerated model selection) partition-model using BIC. Based on a candidate pool of potential partitioning strategies that spanned a single partition for the entire data set to a model allowing each locus to represent a unique partition, the latter approach suggested that 75 partitions of our data set represented the best-fitting partitioning scheme, which reduced the number of necessary model parameters by 71%, and hugely decreased computation time.

We analysed each individual locus in RAxML v8.0.20 (ref. 18), and then the concatenated alignment, using the two partitioning strategies identified above with both maximum likelihood and Bayesian based approaches in RAxML v8.0.20, and ExaBayes v1.4.2 9 (ref. 34). For each RAxML analysis, we executed 100 rapid bootstrap inferences and thereafter a thorough ML search using a GTR+Γ_4 model of nucleotide substitution for each data set partition. Although this may potentially over-parameterize a partition with respect to substitution model, the influence of this form of model over-parameterization has been found to be negligible in phylogenetic inference[35]. For the Bayesian analyses, we ran four Metropolis-coupled ExaBayes replicates for 10 million generations, each with three heated chains, and sampling every 1,000 generations (default tuning and branch swap parameters; branch lengths among partitions were linked). Convergence and proper sampling of the posterior distribution of parameter values were assessed by checking that the effective sample sizes of all estimated parameters and branch lengths were greater than 200 in the Tracer v1.6 software[36] (most were greater than 1,000), and by using the 'sdsf' and 'postProcParam' tools included with the ExaBayes package to ensure the average standard deviation of split frequencies and potential scale reduction factors across runs were close to zero and one, respectively. Finally, to check for convergence in topology and clade posterior probabilities, we summarized a greedily refined majority-rule consensus tree (default) from 10,000 post burn-in trees using the ExaBayes 'consense' tool for each run independently and then together. Analyses of the reduced data set referenced in the main text were conducted using the same partition-model as the full data set.

To explore variation in gene tree topology and to look for outliers that might influence combined analysis, we calculated pairwise Robinson-Foulds[37] (RF) and Matching Splits (MS) tree distances implemented in TreeCmp[38]. We then visualized histograms of tree distances and multidimensional scaling plots in R, and estimated neighbour-joining 'trees-of-trees' in the Phangorn R package sensu lato[39,40]. Using RF and MS distances, outlier loci were identified as those that occurred in the top 10% of pairwise distances for >30 comparisons to other loci (~10%) in the data set. We also identified putative outlier loci using the kdetrees.complete function of the kdetrees R package[41]. All three methods identified 13 of the same loci as potential outliers; however removal of these loci from the analysis had no effect on estimating topology or branch lengths.

Coalescent species tree analyses. Although fully parametric estimation (for example, *BEAST, see ref. 42) of a coalescent species tree with hundreds of genes and hundreds of taxa is not currently possible, we estimated species trees using three gene-tree summation methods that have been shown to be statistically consistent under the multispecies coalescent model[43]. First, we used the STRAW web server[44] to estimate bootstrapped species trees using the STAR[19] and NJ-ST[20] algorithms (also available through STRAW). The popular MP-EST[45] method cannot currently work for more than ~50 taxa. STAR takes rooted gene trees and uses the average ranks of coalescence times[19] to build a distance matrix from which a species tree is computed with the neighbour-joining method[46]. By contrast, NJst applies the neighbour-joining method to a distance matrix computed from average gene-tree internode distances, and relaxes the requirement for input gene trees to be rooted[20].

We also summarized a species tree with the ASTRAL 4.7.6 algorithm. With simulated data, ASTRAL has been shown to outperform concatenation or other summary methods under certain amounts of incomplete lineage sorting[21]. For very large numbers of taxa and genes, ASTRAL uses a heuristic search to find the species tree that agrees with the largest number of quartet trees induced by the set of input gene trees. For analysis with ASTRAL, we also attempted to increase the resolution of individual gene trees (Supplementary Fig. 2) by generating supergene alignments using the weighted statistical binning pipeline of refs 47, 48 with a bootstrap score of 0.75 as a bin threshold.

STAR, NJst (not shown), and the binned ASTRAL (Supplementary Fig. 3) analysis produced virtually identical inferences when low support branches (<0.75) were collapsed, and differed only with respect to the resolution of a few branches. NJst resolved the Passeroidea (*Fringilla* plus *Spizella*) as the sister group to a paraphyletic sample of Sylvioidea (*Calandrella*, *Pycnonotus,* and *Sylvia*), while STAR does not resolve this branch. Comparing STAR/NJst to ASTRAL, we find five additional differences: (1) within tinamous, STAR/NJst resolves *Crypturellus* as sister to the rest of the tinamous, whereas ASTRAL resolves *Crypturellus* as sister to *Tinamus* (similar to ExaBayes/RAxML); (2) STAR/NJst resolves pigeons as sister to a clade containing Mesitornithiformes and Pteroclidiformes, while ASTRAL does not resolve these relationships; (3), STAR/NJst fails to resolve *Oxyruncus* and *Myiobius* as sister genera, while ASTRAL does (similar to RAxML/ExaBayes); (4), in STAR/NJst, bee-eaters (*Merops*) are resolved as the sister group to coraciiforms (congruent with ref. 4), while ASTRAL resolves bee-eaters as sister to the rollers (*Coracias*) (similar to RAxML/ExaBayes); (5) lastly, in STAR/NJst, buttonquail (*Turnix*) is resolved as sister to the most inclusive clade of Charadriiformes not including *Burhinus*, *Charadrius*, *Haematopus*, and *Recurvirostra*, while in ASTRAL, buttonquail is resolved as sister to a clade containing *Glareola*, *Uria*, *Rynchops*, *Sterna*, and *Chroicocephalus* (similar to RAxML/ExaBayes).

Although lower level relationships detected with concatenation are generally recapitulated in the species trees, few of the higher level, or interordinal, relationships are resolved. This lack of resolution of the gene-tree species-tree based inferences relative to the inferences based on concatenation are not surprising, as it is increasingly recognized that the phylogenetic information content required to resolve the gene-tree histories of individual loci becomes scant at deep timescales[47]. Despite our extensive taxon sampling and the slow rate of nucleotide substitution that characterizes loci captured using anchored enrichment[12], no single locus was able to fully resolve a topology, and this lack of information will challenge the accuracy of any coalescent-based summary approach relative to concatenation[49–54]. Finally, all summation methods tested here assume a priori that the only source of discordance among gene trees is deep coalescence, and violations of this assumption may introduce systematic error in phylogeny estimation[54].

Phylogenetic informativeness. Site-specific evolutionary rates, $\lambda_{i...j}$, were calculated for each locus using the program HyPhy[55] in the PhyDesign web interface[56] in conjunction with a guide chronogram generated by a nonparametric rate smoothing algorithm[57] applied to our concatenated RAxML tree. Using these rates to predict whether an alignment will yield correct, incorrect, or no resolution of a given node, we quantified the probability of phylogenetically informative changes

(ψ)[16] contributing to the resolution of the earliest divergences in Neoaves. Estimates generated under a three character state model[58] reveal that the majority of loci have a strong probability of ψ, and suggest a high potential for most loci and partitions containing multiple loci (assigned by PartitionFinder) to correctly resolve this internode. The potential for resolution as a consequence of phylogenetic signal is therefore high relative to the potential for saturation and misleading inference induced by stochastic changes along the subtending lineages (Supplementary Fig. 4a).

To assess the information content of the loci across the entire topology, we profiled their phylogenetic informativeness (PI)[15], (Supplementary Fig. 4b). There was considerable variation in PI across loci (Supplementary Fig. 4). In all cases, the loci with the lowest values of ψ are categorized by substantially lower (60–90%) values of PI, rather than sharp declines in their PI profiles. The absence of a sharp decline in the PI profile suggests that a lack of phylogenetic information, rather than rapid increases in homoplasious sites, underlie low values of the probability of signal ψ[59].

Because declines in PI can be attributed to increases in homoplasious site patterns[59], we further assessed the phylogenetic utility of data set partitions by quantifying the ratio of PI at the most recent common ancestor of Neoaves to the PI at the most recent common ancestor of Aves (Supplementary Fig. 4c). Values of this ratio that are less than 1 correspond to a rise in PI towards the root. Values close to 1 correspond to fairly uniform PI. Values greater than 1 correspond to a decline in PI towards the root. Sixty-six out of 75 partitions demonstrated less than a 50% percent decline in PI, and only six partitions demonstrated a decline of PI greater than 75% (Supplementary Fig. 4c). As all but a few nodes in this study represent divergences younger than the crown of Neoaves, these ratios of PI suggest that the predicted impact of homoplasy on our topological inferences should be minimal.

As PI profiles do not directly predict the impact of homoplasious site patterns on topological resolution[16,60], we evaluated probabilities of ψ for focal nodes using both the concatenated data set as well as individual loci that span the variance in locus lengths. Concordant with expectations from the PI profiles, all quantifications strongly support the prediction that homoplasy will have a minimal impact on topological resolution for the concatenated data set across a range of tree depths and internode distances ($\psi = 1.0$ for all nodes), while individual loci vary in their predicted utility (Supplementary Fig. 4d). As the guide tree does not represent a true known tree, we additionally quantified ψ across a range of tree depths and internode distances to test if our predictions of utility are in line with general trends above. Concordant with our results above, the concatenated data set is predicted to be of high phylogenetic utility at all timescales ($\psi = 1.0$ for all nodes), while the utility of individual loci begins to decline for small internodes at deep tree depths (Supplementary Fig. 5).

Estimating a time-calibrated phylogeny. We estimated a time-calibrated tree with a node dating approach in BEAST 1.8.1 (ref. 42) that used 19 well justified fossil calibrations phylogenetically placed by rigorous, apomorphy-based diagnoses (see the descriptions of avian calibration fossils in the Supplementary Information). We used a starting tree topology based on the ExaBayes inference (Fig. 1), and prior node age calibrations that followed a lognormal parametric distribution based on occurrences of fossil taxa. To prevent BEAST from exploring topology space and only allow estimation of branch lengths, we turned off the subtree-slide, Wilson–Balding, and narrow and wide exchange operators[61,62]. Finally, we applied a birth–death speciation model with default priors.

As rates of molecular evolution are significantly variable across certain bird lineages[63–65], we applied an uncorrelated relaxed clock (UCLN) to each partition of the data set where rates among branches are distributed according to a lognormal distribution[66]. All dating analyses were performed without crocodilian outgroups to reduce the potential of extreme substitution rate heterogeneity to bias rate and consequent divergence time estimates of the UCLN model[67].

All calibrations were modelled using soft maximum age bounds to allow for the potential of our data to overwhelm our user-specified priors[68]. Soft maximum bounds are the preferred method for assigning upper limits on the age of phylogenetic divergences[69]. As effective priors necessarily reflect interactions between user specified priors, topology, and the branching-model, they may not precisely reflect the user-specified priors[70]. To correct for this potential source of error, we carefully examined the effective calibration priors by first running the prepared BEAST XML without any nucleotide data (until all ESS values were above 200). We then iteratively adjusted our user-defined priors until all of the effective priors (as examined in the Tracer software) reflected the intended calibration densities. Finally, using the compare.phylo function in the Phyloch R package, we examined how the inclusion of molecular data influenced the divergence time estimates relative to the effective prior (Supplementary Fig. 9; see below).

Defining priors. Our initial approach was to set a prior's offset to the age of its associated fossil; the mean was then manually adjusted such that 95% of the

calibration density fell more recently than the K–Pg boundary at 65 Ma (million years ago) (the standard deviation was fixed at 1 Ma). In general, priors constructed this way generated calibration densities that specified their highest density peak (their mode) about 3–5 million years older than the age of the offset.

We applied a loose gamma prior to the node reflecting the most recent common ancestor of crown birds—we used an offset of 60.5 Ma (the age of the oldest known definitive, uncontroversial crown bird fossil; the stem penguin *Waimanu*), and adjusted the scale and shape of the prior such that 97.5% of the calibration density fell more recently than 86.5 Ma[71] (see below and Supplementary Information for discussion of the >65 Ma putative crown avian *Vegavis*). This date (86.5 Ma) reflects the upper bound age estimate of the Niobrara Formation—one of many richly fossiliferous Mesozoic deposits exhibiting many crownward Mesozoic stem birds, without any trace of avian crown group representatives. The Niobrara, in particular, has produced hundreds of stem birds and other fragile skeletons, without yielding a single crown bird fossil, and therefore represents a robust choice for a soft upper bound for the root divergence of the avian crown[71–73]. Previous soft maxima employed for this divergence have arbitrarily selected the age of other Mesozoic stem avians (that is, *Gansus yumenensis*, 110 Ma) that are phylogenetically stemward of the Niobrara taxa[28]. Although the implementation of very ancient soft maxima such as the age of *Gansus* are often done in the name of conservatism, the extremely ancient divergence dates yielded by such analyses illustrate the misleading influence of assigning soft maxima that are vastly too old to be of relevance to the divergence of crown group birds[74]. However, this problem has been eliminated in some more recent analyses[75].

All of the fossil calibrations employed in our analysis represent neognaths; rootward divergences within Aves (for example the divergence between Palaeognathae and Neognathae, and Galloanserae and Neoaves) cannot be confidently calibrated due to a present lack of fossils representing the palaeognath, neognath, galloanserine, and neoavian stem groups. As such, the K–Pg soft bound was only applied to comparatively apical divergences within neognaths. Although the question of whether major neognath divergences occurred during the Mesozoic has been the source of controversy[76–78], renewed surveys of Mesozoic sediments for definitive crown avians or even possible crown neoavians have been unsuccessful (with the possible exception of *Vegavis*; see Supplementary Information), and together with recent divergence dating analyses have cast doubt on the presence of neoavian subclades before the K–Pg mass extinction[1,74,79]. Further, recent work has demonstrated the tendency of avian divergence estimates to greatly exceed uninformative priors, resulting in spuriously ancient divergence dating results (for example, refs 28, 75, 76, 80). These results motivated our implementation of the 65 Ma soft bound for our neoavian calibrations.

Contrary to expectation, when we compared the effective prior on the entire tree to the final summary derived from the posterior distribution of divergence times (Supplementary Fig. 9), we found no overall trend of posterior estimated ages post-dating prior calibrations. In fact, the inclusion of our molecular data decreases the inferred ages of almost all of the deepest nodes in our tree. A similar result has been obtained for mammals by using large amounts of nuclear DNA sequences[81]. Future work investigating the interplay of the density of genomic sampling and the application of various calibration age priors will be indispensible for sensitivity analyses to help us further develop a robust timescale of avian evolution. However, the pattern of posterior versus prior age estimates observed in our study raises the prospect that the new class of data used in this study (that is, semi-conserved anchor regions) may exhibit some immunity to longstanding problems associated with inferring avian divergence times, such as systematically over-estimating the antiquity of extant avian clades.

Implementing BEAST and summarizing a final calibrated tree. In addition to making predictions about the phylogenetic utility of a locus or partition towards topological resolution, PI profiles have recently also been used to mitigate the influence of substitution saturation on divergence time estimates[82]. Given the variance in PI profile shapes for captured loci and their subsequent partition assignments (Supplementary Fig. 4c), and observations that alignments and subsets of data alignments characterized by high levels of homoplasy can mislead branch length estimation[83,84], we limited our divergence time estimates to 36 partitions that did not exhibit a decline in informativeness towards the root of the tree. We ran BEAST on each partition separately until parameter ESS values were greater than 200 (most were greater than 1,000) to ensure adequate posterior sampling of each parameter value. After concatenating 10,000 randomly sampled post burn-in trees from each of these completed analyses, we summarized a final MCC tree with median node heights in TreeAnnotator v1.8.1 (ref. 42). Supplementary Fig. 6 shows the full, calibrated Bayesian tree (Fig. 1) with 95% HPD confidence intervals on the node ages, and Supplementary Fig. 7 shows the distribution of estimated branching times, ranked by median age (using clade numbers from Fig. 1). All computations were carried out on 64-core PowerEdge

M915 nodes on the Louise Linux cluster at the Yale University Biomedical High Performance Computing Center.

Data reporting. No statistical methods were used to predetermine sample size.

31. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
32. Meyer, M. & Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* http://dx.doi.org/10.1101/pdb.prot5448 (2010).
33. Rokyta, D. R., Lemmon, A. R., Margres, M. J. & Arnow, K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* **13**, 312 (2012).
34. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
35. Dornburg, A., Santini, F. & Alfaro, M. E. The influence of model averaging on clade posteriors: an example using the triggerfishes (Family Balistidae). *Syst. Biol.* **57**, 905–919 (2008).
36. Tracer. v1.6 http://beast.bio.ed.ac.uk/Tracer (2014).
37. Robinson, D. F. & Foulds, L. R. in *Combinatorial Mathematics VI* in *Lecture Notes in Mathematics*, Vol. 748 (eds Horadam A. F. & Wallis W. D.) Ch. 12 119–126 (Springer, 1979).
38. Bogdanowicz, D., Giaro, K. & Wróbel, B. TreeCmp: comparison of trees in polynomial time. *Evol. Bioinform.* **8**, 475–487 (2012).
39. Nye, T. M. W. Trees of Trees: an approach to comparing multiple alternative phylogenies. *Syst. Biol.* **57**, 785–794 (2008).
40. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
41. Weyenberg, G., Huggins, P. M., Schardl, C. L., Howe, D. K. & Yoshida, R. KDETREES: non-parametric estimation of phylogenetic tree distributions. *Bioinformatics* **30**, 2280–2287 (2014).
42. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
43. Rannala, B. & Yang, Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656 (2003).
44. Shaw, T. I., Ruan, Z., Glenn, T. C. & Liu, L. STRAW: species tree analysis web server. *Nucleic Acids Res.* **41**, W238–W241 (2013).
45. Liu, L., Yu, L. & Edwards, S. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302 (2010).
46. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
47. Mirarab, S., Bayzid, M. S., Boussau, B. & Warnow, T. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* **346** (2014).
48. Mirarab, S., Bayzid, M. S. & Warnow, T. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Syst. Biol.*, (2014).
49. Bayzid, M. S. & Warnow, T. Naive binning improves phylogenomic analyses. *Bioinformatics* **29**, 2277–2284 (2013).
50. DeGiorgio, M. & Degnan, J. H. Fast and consistent estimation of species trees using supermatrix rooted triples. *Mol. Biol. Evol.* **27**, 552–569 (2010).
51. Kimball, R. T., Wang, N., Heimer-McGinn, V., Ferguson, C. & Braun, E. L. Identifying localized biases in large datasets: a case study using the avian tree of life. *Mol. Phylogenet. Evol.* **69**, 1021–1032 (2013).
52. McCormack, J. E. *et al.* A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE* **8**, e54848 (2013).
53. Springer, M. S. & Gatesy, J. Land plant origins and coalescence confusion. *Trends Plant Sci.* **19**, 267–269 (2014).
54. Tonini J., Moore A., Stearn D., Shcheglovitova M. & Ortí, G. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLOS Currents Tree of Life* **1** (2015).
55. Pond, S. L. K. & Muse, S. V. in *Statistical Methods in Molecular Evolution* (ed. Nielsen, R.) 125–181 (Springer, 2005).
56. López-Giráldez, F. & Townsend, J. P. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evol. Biol.* **11**, 152 (2011).
57. Sanderson, M. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**, 1218 (1997).
58. Simmons, M. P., Carr, T. G. & O'Neill, K. Relative character-state space, amount of potential phylogenetic information, and heterogeneity of nucleotide and amino acid characters. *Mol. Phylogenet. Evol.* **32**, 913–926 (2004).
59. Townsend, J. P. & Leuenberger, C. Taxon sampling and the optimal rates of evolution for phylogenetic inference. *Syst. Biol.* **60**, 358–365 (2011).
60. Klopfstein, S., Kropf, C. & Quicke, D. L. J. An evaluation of phylogenetic informativeness profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). *Syst. Biol.* **59**, 226–241 (2010).
61. Drummond, A. J. & Bouckaret, R. R. *Bayesian Evolutionary Analysis With BEAST* (Cambridge Univ. Press, 2015).
62. Hsiang, A. Y. *et al.* The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. *BMC Evol. Biol.* **15**, 87 (2015).
63. Phillips, M. J., Gibb, G. C., Crimp, E. A. & Penny, D. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Syst. Biol.* **59**, 90–107 (2010).
64. Pereira, S. L. & Baker, A. J. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol. Biol. Evol.* **23**, 1731–1740 (2006).

65. Nam, K. *et al.* Molecular evolution of genes in avian genomes. *Genome Biol.* **11,** R68 (2010).
66. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4,** e88 (2006).
67. Dornburg, A., *et al.* Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Mol. Biol. Evol.* **29,** 721–736 (2012).
68. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23,** 212–226 (2006).
69. Ho, S. Y. W. Calibrating molecular estimates of substitution rates and divergence times in birds. *J. Avian Biol.* **38,** 409–414 (2007).
70. Heled, J. & Drummond, A. J. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst. Biol.* **61,** 138–149 (2012).
71. Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24,** 26 (2007).
72. Clarke, J. A. Morphology, phylogenetic taxonomy, and systematics of *Ichthyornis* and *Apatornis* (Avialae: Ornithurae). *Bull. Am. Mus. Nat. Hist.* **286,** 1–179 (2004).
73. Field, D. J., LeBlanc, A., Gau, A. & Behlke, A. D. B. Pelagic neonatal fossils support viviparity and precocial life history of Cretaceous mosasaurs. *Palaeontology* **58,** 401–407 (2015).
74. Mayr, G. The age of the crown group of passerine birds and its evolutionary significance — molecular calibrations versus the fossil record. *Syst. Biodivers.* **11,** 7–13 (2013).
75. Jetz, W. *et al.* Global distribution and conservation of evolutionary distinctness in birds. *Curr. Biol.* **24,** 919–930 (2014).
76. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381,** 226–229 (1996).
77. Benton, M. J. Early origins of modern birds and mammals: molecules vs. morphology. *Bioessays* **21,** 1043–1051 (1999).
78. Hope, S. in *Mesozoic Birds: Above the Heads of Dinosaurs* (eds Chiappe L. M. & Witmer L. M.) 339–388 (Univ. of California Press, 2002).
79. Longrich, N. R., Tokaryk, T. & Field, D. J. Mass extinction of birds at the Cretaceous–Paleogene (K–Pg) boundary. *Proc. Natl Acad. Sci. USA* **108,** 15253–15257 (2011).
80. Baker, A. J., Pereira, S. L. & Paton, T. A. Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. *Biol. Lett.* **3,** 205–209 (2007).
81. dos Reis, M. *et al.* Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B* **279,** 3491–3500 (2012).
82. Dornburg, A., Townsend, J. P., Friedman, M. & Near, T. J. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evol. Biol.* **14,** 169 (2014).
83. Brandley, M. C. *et al.* Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of Plestiodon (Eumeces) lizards. *Syst. Biol.* **60,** 3–15 (2011).
84. Phillips, M. J. Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene* **441,** 132–140 (2009).

LETTER

Biodiversity increases the resistance of ecosystem productivity to climate extremes

Forest Isbell[1], Dylan Craven[2,3], John Connolly[4], Michel Loreau[5], Bernhard Schmid[6], Carl Beierkuhnlein[7], T. Martijn Bezemer[8], Catherine Bonin[9], Helge Bruelheide[2,10], Enrica de Luca[6], Anne Ebeling[11], John N. Griffin[12], Qinfeng Guo[13], Yann Hautier[14], Andy Hector[15], Anke Jentsch[16], Jürgen Kreyling[17], Vojtěch Lanta[18], Pete Manning[19], Sebastian T. Meyer[20], Akira S. Mori[21], Shahid Naeem[22], Pascal A. Niklaus[6], H. Wayne Polley[23], Peter B. Reich[24,25], Christiane Roscher[2,26], Eric W. Seabloom[1], Melinda D. Smith[27], Madhav P. Thakur[2,3], David Tilman[1,28], Benjamin F. Tracy[29], Wim H. van der Putten[8,30], Jasper van Ruijven[31], Alexandra Weigelt[2,3], Wolfgang W. Weisser[20], Brian Wilsey[32] & Nico Eisenhauer[2,3]

It remains unclear whether biodiversity buffers ecosystems against climate extremes, which are becoming increasingly frequent worldwide[1]. Early results suggested that the ecosystem productivity of diverse grassland plant communities was more resistant, changing less during drought, and more resilient, recovering more quickly after drought, than that of depauperate communities[2]. However, subsequent experimental tests produced mixed results[3–13]. Here we use data from 46 experiments that manipulated grassland plant diversity to test whether biodiversity provides resistance during and resilience after climate events. We show that biodiversity increased ecosystem resistance for a broad range of climate events, including wet or dry, moderate or extreme, and brief or prolonged events. Across all studies and climate events, the productivity of low-diversity communities with one or two species changed by approximately 50% during climate events, whereas that of high-diversity communities with 16–32 species was more resistant, changing by only approximately 25%. By a year after each climate event, ecosystem productivity had often fully recovered, or over-shot, normal levels of productivity in both high- and low-diversity communities, leading to no detectable dependence of ecosystem resilience on biodiversity. Our results suggest that biodiversity mainly stabilizes ecosystem productivity, and productivity-dependent ecosystem services, by increasing resistance to climate events. Anthropogenic environmental changes that drive biodiversity loss thus seem likely to decrease ecosystem stability[14], and restoration of biodiversity to increase it, mainly by changing the resistance of ecosystem productivity to climate events.

Biodiversity stabilizes ecosystem productivity over time[9,14–23]; however, it remains unclear whether it does so by providing resistance during climate events, resilience (*sensu* rapid recovery[24]) after climate events, or both (Extended Data Fig. 1). Two decades ago, a seminal study reported that the ecosystem productivity of diverse grassland plant communities was more resistant and more resilient to a major drought than that of depauperate communities[2]. However, this study had not experimentally manipulated biodiversity, which confounded variation in biodiversity with variation in species composition and resource availability[25]. Hundreds of biodiversity experiments were subsequently conducted[26,27], but few of these studies revisited this important question, and those that did so found mixed results[3–13]. Further analysis of the original data also produced mixed results[28]. Thus, it remains unclear whether biodiversity buffers ecosystems against climate extremes, which are becoming increasing frequent worldwide[1].

We combined data from 46 experiments that manipulated grassland plant diversity and measured productivity across Europe and North America (Extended Data Fig. 2 and Extended Data Table 1). We classified each year of each experiment as extremely dry, moderately dry, normal, moderately wet, or extremely wet (Extended Data Figs 2 and 3) (Methods). To do this in a globally consistent manner, we used a drought index that quantified month-by-month variations in water balance over the past century on 0.5 degree × 0.5 degree grids globally, based on measurements at more than 4,000 weather stations worldwide[29,30] (Extended Data Figs 2 and 3). We defined climate extremes (extremely dry or extremely wet) as events occurring less frequently than once per decade, based on the historic climate at each site over the past century (Methods). Moderately dry and wet events were defined as those that had historically occurred between once in 4 years and once per decade. Normal years included the interquartile range of observed water balances. Given these cutoffs, there were 18 extremely dry, 32 moderately dry, 87 normal, 37 moderately wet, and 21 extremely wet experiment years that occurred during these biodiversity experiments (Extended Data Figs 2 and 3). Unsurprisingly, productivity tended to be lower than normal during dry events and higher than normal during wet events (Extended Data Fig. 4), although

[1]Department of Ecology, Evolution and Behavior, University of Minnesota Twin Cities, Saint Paul, Minnesota 55108, USA. [2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. [3]Institute of Biology, Leipzig University, Johannisallee 21, 04103 Leipzig, Germany. [4]Ecological and Environmental Modelling Group, School of Mathematics and Statistics, University College Dublin, Dublin 4, Ireland. [5]Centre for Biodiversity Theory and Modelling, Experimental Ecology Station, Centre National de la Recherche Scientifique, Moulis 09200, France. [6]Institute of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland. [7]Department of Biogeography, BayCEER, University of Bayreuth, 95440 Bayreuth, Germany. [8]Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), PO Box 50, 6700 AB Wageningen, The Netherlands. [9]Department of Agronomy, Iowa State University, Ames, Iowa 50011, USA. [10]Institute of Biology, Martin Luther University Halle-Wittenberg, 06108 Halle, Germany. [11]Institute of Ecology, Friedrich Schiller University Jena, Dornburger Strasse 159, 07743 Jena, Germany. [12]Department of Biosciences, Swansea University, Singleton Park, Swansea SA28PP, UK. [13]USDA FS, Eastern Forest Environmental Threat Assessment Center, RTP, North Carolina 27709, USA. [14]Ecology and Biodiversity Group, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. [15]Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. [16]Disturbance Ecology, BayCEER, University of Bayreuth, 95440 Bayreuth, Germany. [17]Institute of Botany and Landscape Ecology, Ernst-Moritz-Arndt University Greifswald, D-17487 Greifswald, Germany. [18]Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 31, 37005 České Budějovice, Czech Republic. [19]Institute for Plant Sciences, University of Bern, CH-3013 Bern, Switzerland. [20]Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Technische Universität München, 85354 Freising, Germany. [21]Graduate School of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, Hodogaya, Yokohama, Kanagawa, 240-8501, Japan. [22]Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, New York 10027, USA. [23]US Department of Agriculture Agricultural Research Service, Grassland, Soil and Water Research Laboratory, Temple, Texas 76502, USA. [24]Department of Forest Resources, University of Minnesota Twin Cities, Saint Paul, Minnesota 55108 USA. [25]Hawkesbury Institute for the Environment, University of Western Sydney, Penrith, New South Wales 2753, Australia. [26]UFZ Helmholtz Centre for Environmental Research, Community Ecology, 06120 Halle, Germany. [27]Graduate Degree Program in Ecology and Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA. [28]Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106 USA. [29]Crop and Soil Environmental Sciences, Smyth Hall 0404, Virginia Tech, Blacksburg, Virginia 24061, USA. [30]Laboratory of Nematology, Wageningen University and Research Centre, PO Box 8123, 6700 ES Wageningen, The Netherlands. [31]Nature Conservation and Plant Ecology Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands. [32]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA.

Table 1 | Fixed effect tests and variance component estimates (standard error) for linear mixed-effects models

	Resistance	Resilience
Fixed effects		
Biodiversity	$F_{1,27.8} = 20.68$***	$F_{1,8.5} = 0.67$
Direction	$F_{1,81.7} = 0.53$	$F_{1,56.9} = 0.15$
Intensity	$F_{1,85.6} = 1.40$	$F_{1,57.7} = 2.36$
Biodiversity × intensity	$F_{1,82.3} = 3.02$*	
Biodiversity × direction		$F_{1,46.1} = 6.52$**
Variance components		
Study	0.37 (0.15)	1.4×10^{-6} (3.5×10^{-8})
Study × biodiversity	0.041 (0.022)	0.0067 (0.0096)
Study × year	0.32 (0.074)	0.68 (0.15)
Study × biodiversity × year	0.033 (0.011)	0.018 (0.012)
Plot	0.25 (0.038)	9.6×10^{-7} (2.3×10^{-8})
Plot × year	2.1 (0.051)	4.1 (0.099)
Temporal autocorrelation		
ρ_{AR1}	0.12 (0.025)	−0.41 (0.020)

*$P < 0.1$; **$P < 0.05$; ***$P < 0.001$. Direction: 0, dry; 1, wet. Intensity: 0, moderate; 1, extreme. Biodiversity: $\log_2$(number of species). Study = factor. Year = factor. Plot is defined within studies. Both response variables were $\log_2$-transformed. Non-significant ($P > 0.1$) interactions were excluded from the model. Kenward–Roger approximation is given for denominator degrees of freedom.

there were exceptions to this general trend (Extended Data Fig. 5). Productivity overshot normal levels when recovering during the year after extreme (but not moderate) dry and wet events (Extended Data Fig. 4), which is consistent with damped oscillations, rather than monotonic recovery, of productivity after climate extremes (Extended Data Fig. 1). Consistent with previous studies[9,14–23], biodiversity increased ecosystem stability (Fig. 1a; $F_{1,37.4} = 28.74$, $P < 0.001$).

We quantified resistance and resilience, using proportional changes in productivity from one year to the next, within each experimental unit (plot) for each observed climate event (Methods). Linear mixed-effects models were used to test whether resistance and resilience depend on biodiversity, and how these biodiversity effects depend on climate event properties, such as the direction (wet or dry), intensity (moderate or extreme), or duration (3–24 months) of climate events, while accounting for repeated measurements (Methods).

Biodiversity increased the resistance of ecosystem productivity to a broad range of climate events (biodiversity main effect in Table 1 and Fig. 1b). That is, more diverse communities exhibited smaller proportional changes in productivity during climate events. On average, across all studies and climate events, the productivity of low-diversity

communities with one or two species changed by approximately 50% ($\Omega \approx 2$; Fig. 1b), whereas that of high-diversity communities with 16–32 species changed by approximately 25% ($\Omega \approx 4$; Fig. 1b), during climate events. Biodiversity increased resistance irrespective of the direction (wet or dry) or intensity (moderate or extreme) of climate events (all interactions were non-significant, $P > 0.05$; Table 1). There was, however, one marginally significant interaction: biodiversity may have increased resistance more during moderate climate events than during extreme ones (biodiversity × intensity interaction in Table 1 and Extended Data Fig. 6). There was substantial variability in the effect of biodiversity on resistance among studies and among years within studies (see variance components in Table 1, Fig. 1b and Extended Data Fig. 7); however, biodiversity increased resistance similarly in long-term studies that were conducted for at least 9 years, and in short-term studies (Methods).

Examination of the dynamics of recovery shows that, at both low and high diversity, productivity had often returned to, or overshot, its normal level during the year after a climate event (Extended Data Fig. 4). Given this rapidity of recovery both for low- and for high-diversity communities, biodiversity may not have a major impact on the recovery of ecosystem productivity after climate events, at least over the timescales and climate-event intensities considered. Indeed, we were unable to detect strong and consistent effects of biodiversity on our measure of ecosystem resilience (Table 1 and Fig. 1c). Biodiversity decreased resilience after wet events, and increased, although non-significantly (see confidence intervals for 12-month events shown in Fig. 2), resilience after dry events (biodiversity × direction interaction in Table 1 and Fig. 1c). That is, less diverse communities recovered closer to normal levels of productivity during the year after wet events. On average, across all studies, climate events, and levels of biodiversity, productivity moved approximately 10% closer to normal levels ($\Delta \approx 1.1$; Fig. 1c) during the year after climate events; however, this was often due to greatly overshooting, rather than failing to reach, normal levels of productivity (Extended Data Fig. 4). The effect of biodiversity on resilience did not vary substantially among studies or among years within studies (see relatively small point estimates with large standard errors for biodiversity variance components in Table 1 and Extended Data Fig. 8).

Next, we tested how our results depended on the duration over which climate events were defined. To do so, we considered multiple
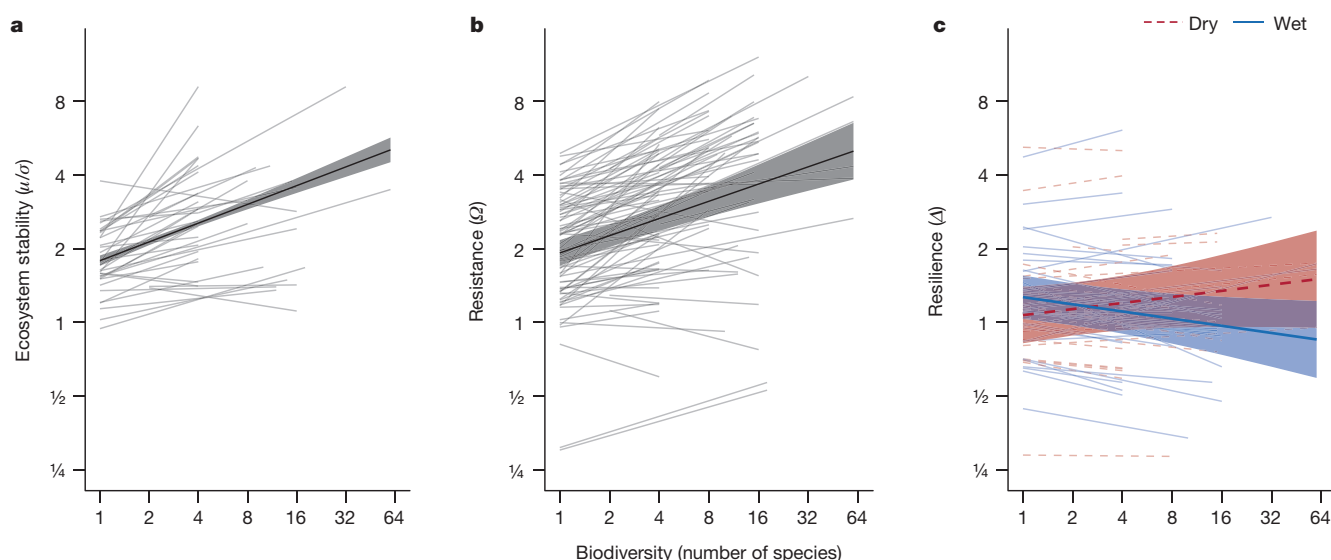

Figure 1 | Biodiversity effects on ecosystem stability, and its resistance and resilience components. Biodiversity consistently increases ecosystem stability (**a**) and resistance (**b**), but not resilience (**c**). Lines are mixed-effects model fits for each study (**a**), or each climate event within each study (**b, c**) (thin lines), or across climate events and studies (thick lines with bands indicating 95% confidence intervals). Thick lines and bands in **c** indicate trends averaged across both moderate and extreme events for either dry (dashed red lines) or wet (solid blue lines) events. Stability measures are unitless. Axes are logarithmic. See Table 1 for test statistics and Extended Data Table 1 for sample sizes.
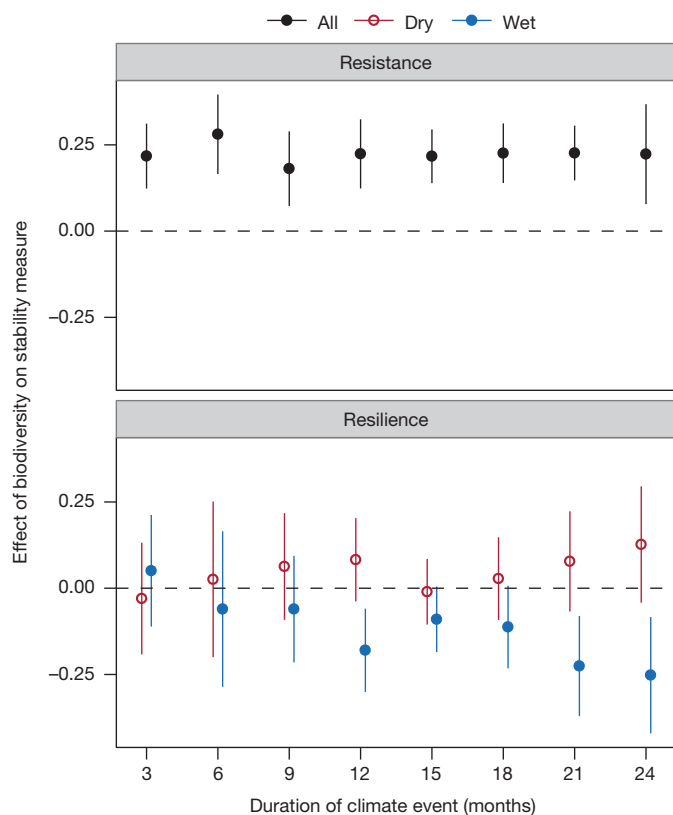
Figure 2 | Effects of biodiversity on stability measures with climate events defined over shorter or longer durations. Biodiversity consistently increases resistance; however, biodiversity effects on resilience depend on the direction (wet or dry) and duration of climate events. Values shown are parameter estimates and 95% confidence intervals for biodiversity effects from mixed-effects models, with the 12-month values corresponding to the results shown in Table 1 and Fig. 1. Values in the upper panel are averaged across both intensities and both directions. For clarity, values in the lower panel are slightly offset on the x axis. See Extended Data Table 1 for sample sizes.

versions of the drought index, which aggregated water balances over different timescales, ranging from seasonal (3 months) to multi-year (24 months) events[30] (Methods). We found that biodiversity consistently increased the resistance of ecosystem productivity during climate events, irrespective of the duration (3–24 months) of the climate event (Fig. 2). Biodiversity had no significant effect on the resilience of ecosystem productivity after brief, intra-annual wet or dry climate events (Fig. 2). Biodiversity decreased resilience only after prolonged, wet climate events that lasted 1 year or more (Fig. 2). The magnitudes of biodiversity effects on resistance were substantially larger than those on resilience for all but the longest durations (Fig. 2).

It is difficult, or perhaps impossible, to fully disentangle the resistance and resilience components of empirical time series, especially when there are frequent perturbations. For example, resilience to the first of two consecutive climate events could bias estimates of resistance to the second event. Similarly, resistance to the second of two consecutive climate events could bias estimates of resilience to the first event. To explore how this might have affected our results, we tested whether biodiversity effects on resistance differed between climate events that were preceded either by normal or by other climate event years, and whether biodiversity effects on resilience differed between climate events that were succeeded either by normal or by climate event years (Methods). We found that biodiversity increased resistance, especially during climate events that were preceded by climate event years (biodiversity × consecutive interaction: $F_{1,64.8} = 7.21$, $P < 0.01$) (Extended Data Fig. 9), and that biodiversity did not significantly impact resilience, regardless of whether a climate event was succeeded
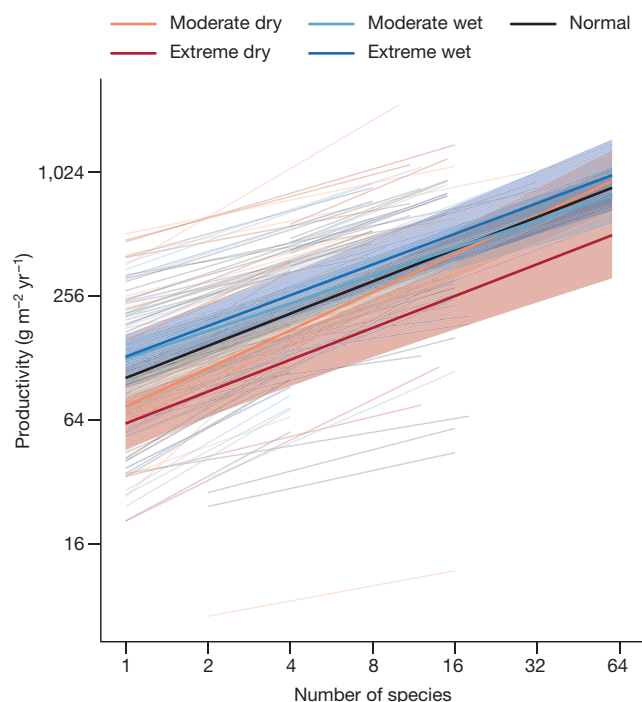
Figure 3 | Biodiversity effects on productivity during climate events or normal years. Lines are mixed-effects model fits for each year within each study (thin lines) or across all years and studies (thick lines with bands indicating 95% confidence intervals). See Extended Data Fig. 5 for results within studies. There was a significant effect of biodiversity on productivity ($F_{1,30.6} = 202.4$, $P < 0.001$), a significant effect of event ($F_{4,139.5} = 6.86$, $P < 0.001$), and a significant biodiversity × event interaction ($F_{4,124.3} = 3.23$, $P = 0.015$). Axes are logarithmic. See Extended Data Table 1 for sample sizes.

by a normal year or another climate event (biodiversity × consecutive interaction: $F_{1,39.6} = 2.42$, $P = 0.13$). We also tested whether biodiversity significantly influenced resilience when considering only climate events that were succeeded by multiple normal years in long-term studies that were conducted for at least 9 years, and with resilience quantified 2, rather than 1, years after climate events (Methods). We again found no detectable effect of biodiversity on resilience ($F_{1,10.6} = 0.20$, $P = 0.66$). Thus, biodiversity did not influence resilience after 1 or 2 years of unperturbed recovery.

Our results suggest that greater biodiversity generally provides greater resistance. We focused on dimensionless, proportional measures of resistance and resilience to allow comparisons of communities with different levels of productivity. However, absolute measures of resistance and resilience might be of interest for some applications within particular communities, and do not necessarily depend on biodiversity in the same manner (Fig. 3 and Extended Data Figs 4 and 5). Given that biodiversity increases productivity, more productivity could be lost during dry events, and gained back after dry events, in diverse than in depauperate communities[3,10]. In this case, it is also important to note that our analyses show that biodiversity increased productivity not only during normal years, but also during climate events (Fig. 3).

Our results suggest that biodiversity stabilizes ecosystem productivity, and probably productivity-dependent ecosystem services, during climate events that are moderate or extreme. Anthropogenic environmental changes that drive biodiversity loss will probably decrease ecosystem stability[14] by decreasing the resistance of ecosystem productivity to climate events. Restoring biodiversity will probably increase ecosystem resistance to climate extremes, which are forecast to become increasingly frequent as the global climate continues to change.

1. Stocker, T. F. *et al.* IPCC 2013: Summary for Policy Makers (Cambridge Univ. Press, 2013).
2. Tilman, D. & Downing, J. A. Biodiversity and stability in grasslands. *Nature* **367,** 363–365 (1994).
3. Pfisterer, A. B. & Schmid, B. Diversity-dependent production can decrease the stability of ecosystem functioning. *Nature* **416,** 84–86 (2002).
4. Caldeira, M. C., Hector, A., Loreau, M. & Pereira, J. S. Species richness, temporal variability and resistance of biomass production in a Mediterranean grassland. *Oikos* **110,** 115–123 (2005).
5. Kahmen, A., Perner, J. & Buchmann, N. Diversity-dependent productivity in semi-natural grasslands following climate perturbations. *Funct. Ecol.* **19,** 594–601 (2005).
6. Steiner, C. F., Long, Z. T., Krumins, J. A. & Morin, P. J. Population and community resilience in multitrophic communities. *Ecology* **87,** 996–1007 (2006).
7. Wang, Y., Yu, S. & Wang, J. Biomass-dependent susceptibility to drought in experimental grassland communities. *Ecol. Lett.* **10,** 401–410 (2007).
8. Ives, A. R. & Carpenter, S. R. Stability and diversity of ecosystems. *Science* **317,** 58–62 (2007).
9. Griffin, J. N. *et al.* in *Biodiversity, Ecosystem Functioning, & Human Wellbeing: An Ecological and Economic Perspective* (eds Naeem, S. *et al.*) Ch. 6 78–93 (Oxford Univ. Press, 2009).
10. van Ruijven, J. & Berendse, F. Diversity enhances community recovery, but not resistance, after drought. *J. Ecol.* **98,** 81–86 (2010).
11. Vogel, A., Scherer-Lorenzen, M. & Weigelt, A. Grassland resistance and resilience after drought depends on management intensity and species richness. *PLoS One* **7,** e36992 (2012).
12. Grossiord, C. *et al.* Tree diversity does not always improve resistance of forest ecosystems to drought. *Proc. Natl Acad. Sci. USA* **111,** 14812–14815 (2014).
13. Wright, A. J. *et al.* Flooding disturbances increase resource availability and productivity but reduce stability in diverse plant communities. *Nature Commun.* **6,** 6092 (2015).
14. Hautier, Y. *et al.* Plant ecology. Anthropogenic environmental changes affect ecosystem stability via biodiversity. *Science* **348,** 336–340 (2015).
15. McNaughton, S. J. Diversity and stability of ecological communities: a comment on the role of empiricism in ecology. *Am. Nat.* **111,** 515–525 (1977).
16. Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390,** 507–509 (1997).
17. Bai, Y., Han, X., Wu, J., Chen, Z. & Li, L. Ecosystem stability and compensatory effects in the Inner Mongolia grassland. *Nature* **431,** 181–184 (2004).
18. Tilman, D., Reich, P. B. & Knops, J. M. H. Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature* **441,** 629–632 (2006).
19. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9,** 1146–1156 (2006).
20. Hector, A. *et al.* General stabilizing effects of plant diversity on grassland productivity through population asynchrony and overyielding. *Ecology* **91,** 2213–2220 (2010).
21. de Mazancourt, C. *et al.* Predicting ecosystem stability from community composition and biodiversity. *Ecol. Lett.* **16,** 617–625 (2013).
22. Polley, H. W., Isbell, F. I. & Wilsey, B. J. Plant functional traits improve diversity-based predictions of temporal stability of grassland productivity. *Oikos* **122,** 1275–1282 (2013).
23. Gross, K. *et al.* Species richness and the temporal stability of biomass production: a new analysis of recent biodiversity experiments. *Am. Nat.* **183,** 1–12 (2014).
24. Pimm, S. L. The complexity and stability of ecosystems. *Nature* **307,** 321–326 (1984).
25. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110,** 449–460 (1997).
26. Cardinale, B. J. *et al.* The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98,** 572–592 (2011).
27. Tilman, D., Isbell, F. & Cowles, J. M. Biodiversity and ecosystem functioning. *Annu. Rev. Ecol. Evol. Syst.* **45,** 471–493 (2014).
28. Tilman, D. Biodiversity: population versus ecosystem stability. *Ecology* **77,** 350–363 (1996).
29. Beguería, S., Vicente-Serrano, S. M. & Angulo-Martinez, M. A multiscalar global drought dataset: the SPEIbase: a new gridded product for the analysis of drought variability and impacts. *Bull. Am. Meteorol. Soc.* **91,** 1351–1356 (2010).
30. Vicente-Serrano, S. M. *et al.* Response of vegetation to drought time-scales across global land biomes. *Proc. Natl Acad. Sci. USA* **110,** 52–57 (2013).

METHODS

Defining ecosystem stability measures. We define measures of resistance and resilience that are (1) dimensionless, and thus directly comparable between studies and communities with different levels of productivity; (2) symmetric, and thus directly comparable between positive and negative perturbations, such as wet and dry climate events; (3) applicable to dynamic systems that exhibit either monotonic recovery or damped oscillations after a perturbation (Extended Data Fig. 1). We define resistance as

$$\Omega \equiv \frac{\overline{Y_n}}{\left| Y_e - \overline{Y_n} \right|} \quad (1)$$

and resilience as

$$\Delta \equiv \left| \frac{Y_e - \overline{Y_n}}{Y_{e+1} - \overline{Y_n}} \right| \quad (2)$$

where $\overline{Y_n}$, Y_e, and Y_{e+1} are respectively the expected ecosystem productivity during normal years (mean across all non-climate event years), during a climate event, and during the year after a climate event. Resistance indicates the proximity of productivity to normal levels during a climate event. For example, if productivity is reduced during a drought to half its normal level, then $\Omega = 2$ (Extended Data Fig. 1). Resilience indicates the rate of return towards normal productivity levels after a climate event. If a climate event lowers productivity, greater biomass growth rates during recovery lead to greater resilience up until they are sufficiently rapid to lead to full recovery of normal levels of productivity during the subsequent year. Any biomass growth rates greater than this lead to progressively lower resilience because productivity overshoots its normal level. Thus, consistent with stability measures used in theoretical biodiversity–stability studies, this measure of resilience has a low value, indicating instability, when the deviation of the system from normal productivity levels exponentially decays at a slow rate, either via monotonic recovery or damped oscillations (Extended Data Fig. 1). For example, if during the year after a climate event productivity recovers either from 50 to 75% or from 50 to 125% of normal productivity levels, then productivity will have returned halfway from perturbed to normal levels, and $\Delta = 2$ (Extended Data Fig. 1). The same is true for recovery in the opposite direction after a positive deviation: that is, recovery from 150 to 125% or from 150 to 75% of normal productivity levels would also give $\Delta = 2$ (Extended Data Fig. 1). The points shown in Extended Data Fig. 1 are given by $y_{t=0} = 100$, $y_{t=1} = 100 - 100/\Omega$, $y_{t=11} = 100 + 100/\Omega$, and, for all other t, $y_t = 100 - (100 - y_{t-1})/\Delta$ for monotonic recovery or $y_t = 100 + (100 - y_{t-1})/\Delta$ for damped oscillations, where y is productivity. We use a common measure of ecosystem stability, quantified as the ratio of the mean to the standard deviation of productivity across years (μ/σ). This measure of ecosystem stability is dimensionless, and thus directly comparable between studies and communities with different levels of productivity.

Identifying wet and dry climate events. Drought occurs when water availability remains below normal levels over some period of time[30]. Identifying and quantifying droughts requires consideration of water inputs (precipitation) and water losses (potential evapotranspiration). Furthermore, doing so in a globally consistent manner requires standardization of spatially explicit historical trends for water balances, to ensure that 'normal' and 'extreme' conditions are consistently defined across sites. Finally, given that ecosystems need not similarly respond to brief or prolonged droughts, it is often useful to consider water balances aggregated over a range of short to long timescales.

We used the standardized precipitation–evapotranspiration index (SPEI) to consistently identify and quantify wet and dry climate events across field experiments over durations ranging from 3 to 24 months. SPEI is a standard normal variable for water balances aggregated over a given number of months at a particular location. SPEI values are based on month-by-month variations in climate over the past century (January 1901 to December 2011), based on monthly means of measurements made at more than 4,000 weather stations worldwide, and provided on 0.5 degree × 0.5 degree grids globally. For example, a value of SPEI-12 = −1.28 for August 2005 at a particular location would correspond to a level of annual (as indicated by the value of 12) drought (as indicated by the negative value) that has historically occurred (between 1901 and 2011) once per decade at that location during the months of September to August (Extended Data Figs 2 and 3). Similarly, SPEI-3 = 0.67 for August 2005 at a particular site would correspond to a level of seasonal wetness that has historically occurred once every 4 years at that particular site during the months of June to August (Extended Data Figs 2 and 3).

We extracted SPEI values from SPEIbase[29] raster files for each peak biomass harvest at each study site (Extended Data Figs 2 and 3). First, we considered annual water balances: SPEI-12. Previous results suggest that primary productivity responds to approximately annual water balances in temperate grasslands[30]. We

classified experiment years as extremely dry, moderately dry, normal, moderately wet, and extremely wet (Extended Data Figs 2 and 3). Extreme events (extremely dry or extremely wet) were defined as those that historically occurred less frequently than once per decade. Moderate events were defined as those that historically occurred between once in 4 years and once per decade. Normal years were defined as those within the interquartile range of historical water balances. Given these cutoffs, there were 18 extremely dry, 32 moderately dry, 87 normal, 37 moderately wet, and 21 extremely wet experiment years that occurred during these biodiversity experiments (Extended Data Figs 2 and 3). Thus, 20% of the experiment years (18 + 21 = 39 out of 195) were identified as extreme events, which corresponds to extremely dry events that occur less than once per decade (10% of observations) plus extremely wet events that occur less than once per decade (10% of observations). Note that there is an unavoidable shifting baseline for comparisons when defining extreme climate events. If we had defined climate extremes based only on data from the early (or late) 1900s, then we would probably have identified more (or fewer) extreme climate events.

Next, we considered how the effects of biodiversity on resistance and resilience depended on the duration over which water balances were aggregated. Specifically, we re-classified each experiment year as extremely dry, moderately dry, normal, moderately wet, and extremely wet years based on other versions of SPEI that aggregate water balances over shorter (SPEI-3, SPEI-6, SPEI-9) or longer (SPEI-15, SPEI-18, SPEI-21, SPEI-24) periods of time preceding peak biomass harvests, and then re-fitted mixed-effects models.

Statistical analyses. We used linear mixed-effects models to test whether resistance and resilience depend on biodiversity, and how these biodiversity effects depend on climate event properties, such as the direction (wet or dry), intensity (moderate or extreme), or duration (3–24 months) of climate events, while accounting for repeated measurements. Models were first fitted with annual (12-month) climate events (Table 1 and Fig. 1), and then subsequently fitted for shorter or longer durations (Fig. 2). Fixed effects were included for biodiversity, quantified as the $\log_2$(treatment species richness); direction, quantified as a binary variable (0, dry; 1, wet); and intensity, quantified as a binary variable (0, moderate; 1, extreme). All interactions were initially included, and non-significant interactions ($P > 0.1$) were subsequently excluded. Random effects were included for a study factor; a study × biodiversity interaction; a study × year interaction; a study × biodiversity × year interaction; and a plot (within-study) term. The error structure accounted for repeated measurements within experimental units (plots) across years. A first-order autoregressive covariance structure provided a better fit than a compound symmetry (split-plot-in-time) covariance structure, according to the Akaike information criterion. For all models, the response variable was $\log_2$-transformed to meet model assumptions.

Models were fitted with the asreml function in the asreml package in R, and results were extracted with the test.asreml function in the pascal package (https://github.com/pascal-niklaus/pascal) in R. After model simplification, as described above on the basis of significance of fixed effects and Akaike information criterion comparisons of random effect and covariance structures, fixed effects were specified as ~biodiversity + direction + intensity + interaction (where interaction = biodiversity:intensity for resistance, and interaction = biodiversity:direction for resilience), random effects as ~study/(biodiversity*year) + plot, and the error structure as rcov = ~id(plot):ar1(year). These mixed-effects models were fitted for annual resistance and resilience (Fig. 1 and Extended Data Figs 7 and 8), and for all eight durations of resistance and resilience (Fig. 2). The model for productivity only differed in the specification of fixed effects, with a factor for climate event (levels of 'extreme dry', 'moderate dry', 'normal', 'moderate wet', and 'extreme wet') instead of the direction and intensity terms (Fig. 3 and Extended Data Figs 4 and 5). The biodiversity × event interaction was significant and retained in the productivity model (Fig. 3).

Models were fitted for resistance for all studies for which there were observations of productivity during both normal and climate event years (Extended Data Figs 3 and 7). Models for resilience were fitted for all studies for which there were observations during normal, climate event, and post-climate event years, except where the only normal year was also the only post-event year because in this case $\overline{Y_n} = Y_{e+1}$ and resilience is undefined (Extended Data Figs 3 and 8).

Species richness treatments were randomly assigned to experimental units (plots). Sample sizes were chosen within individual experiments (Extended Data Table 1) to ensure adequate power to detect an effect of richness on productivity.
Testing whether biodiversity effects differed between short- and long-term studies. Given that many of these studies were conducted for only a few years, we tested whether our results differed between short- and long-term studies. We did so by adding a two-way biodiversity × study duration interaction, and a study duration main effect, to the models shown in Table 1, where study duration was a binary variable with a value of one for the six studies conducted for at least 9 years (Extended Data Table 1), and a value of zero for all other studies. We found

similar results between short- and long-term studies, as indicated by non-significant interactions between biodiversity and study duration for both resistance ($F_{1,16.5} = 0.02$, $P = 0.90$) and resilience ($F_{1,23.7} = 0.66$, $P = 0.42$).

Testing whether biodiversity effects differed between categorical versus continuous measures of climate event intensity. We used a categorical specification of climate intensity (moderate or extreme) throughout because there were often complex nonlinear relationships between biomass production and SPEI within sites (Extended Data Fig. 5). However, our categorical specification incurs some information loss, so we also tested whether results were similar when the models shown in Table 1 were fitted using the absolute value of the SPEI-12 index in place of the binary intensity variable. We found similar results when we considered this continuous measure of climate event intensity. That is, biodiversity increased resistance ($F_{1,28.0} = 20.38$, $P < 0.001$) and did not affect resilience ($F_{1,8.5} = 0.66$, $P = 0.44$).

Disentangling resistance and resilience. It is difficult, or perhaps impossible, to fully disentangle the resistance and resilience components of empirical time series, especially when there are frequent perturbations. For example, resilience to the first of two consecutive climate events could bias estimates of resistance to the second event, and resistance to the second of two consecutive climate events could bias estimates of resilience to the first event. To explore how this affected our results, we added a two-way biodiversity $\times$ consecutive interaction to the models shown in Table 1, and a main effect of consecutive, where consecutive was a binary variable with a value of 1 indicating non-consecutive climate events (that is, normal year before event for resistance, normal year after event for resilience), and 0 otherwise. We also tested whether biodiversity significantly influenced resilience when considering only climate events that were succeeded by multiple normal years in long-term studies that were conducted for at least 9 years, and with resilience quantified 2, rather than 1, years after climate events. To do so, we re-fitted the model shown in Table 1, but with resilience quantified using Y_{e+2} rather than Y_{e+1} in equation (2).

Robustness of results to monoculture exclusion. Given that monocultures are rare in nature, we tested whether our results depended on inclusion of monoculture plots. We found similar results when we excluded monocultures. That is, biodiversity increased resistance and did not significantly affect resilience when we refitted the models shown in Table 1 after excluding monocultures (biodiversity effect on resistance: $F_{1,20.2} = 7.25$, $P = 0.014$; biodiversity effect on resilience: $F_{1,4.4} = 0.21$, $P = 0.665$).
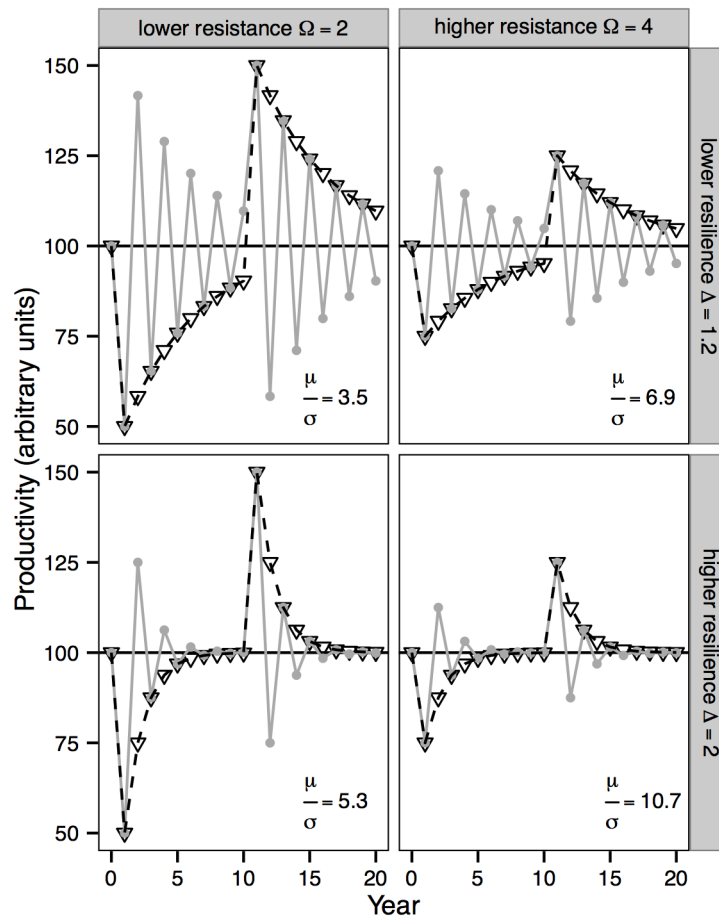
Extended Data Figure 1 | Contrasting ecosystem productivity responses to climate events for low or high levels of resistance (Ω) and resilience (Δ). In these stylized examples, productivity is decreased by a dry climate event during year one, is increased by a wet climate event during year 11, and is otherwise recovering back towards normal productivity levels either monotonically (black dashed lines and open triangles) or via damped oscillations (solid grey lines and filled circles). Ecosystem stability (μ/σ) depends on both resistance and resilience. See Methods for definitions of resistance and resilience.
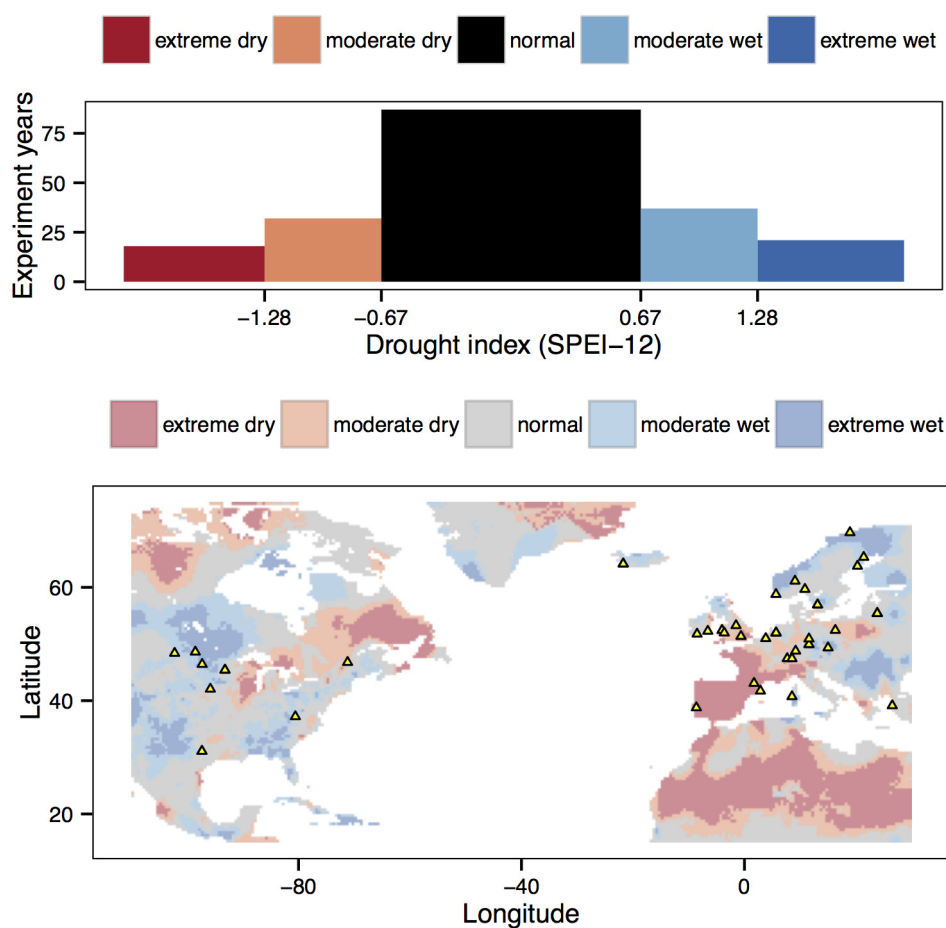
Extended Data Figure 2 | Map of study site locations (bottom) and frequency of climate events (top). Bottom: locations for all 46 studies (yellow triangles) and an example of spatial variation in water balance, where SPEI-12 was classified as in the bottom panel. August 2005 was chosen for this example because many experiments were underway and harvested during this particular month of this particular year (Extended Data Table 1). The spatial patterns of wet and dry climate events shown on this map would differ at other times (that is, during a different month or year) and for climate events defined over other durations (that is, based on water balances aggregated over more or fewer than the preceding 12 months). There were multiple experiments at some sites (Extended Data Table 1), thus some symbols completely overlap on this map. Top: cutoffs for bins correspond to events occurring every 1 in 4 years (± 0.67) or every 1 in 10 years (± 1.28).
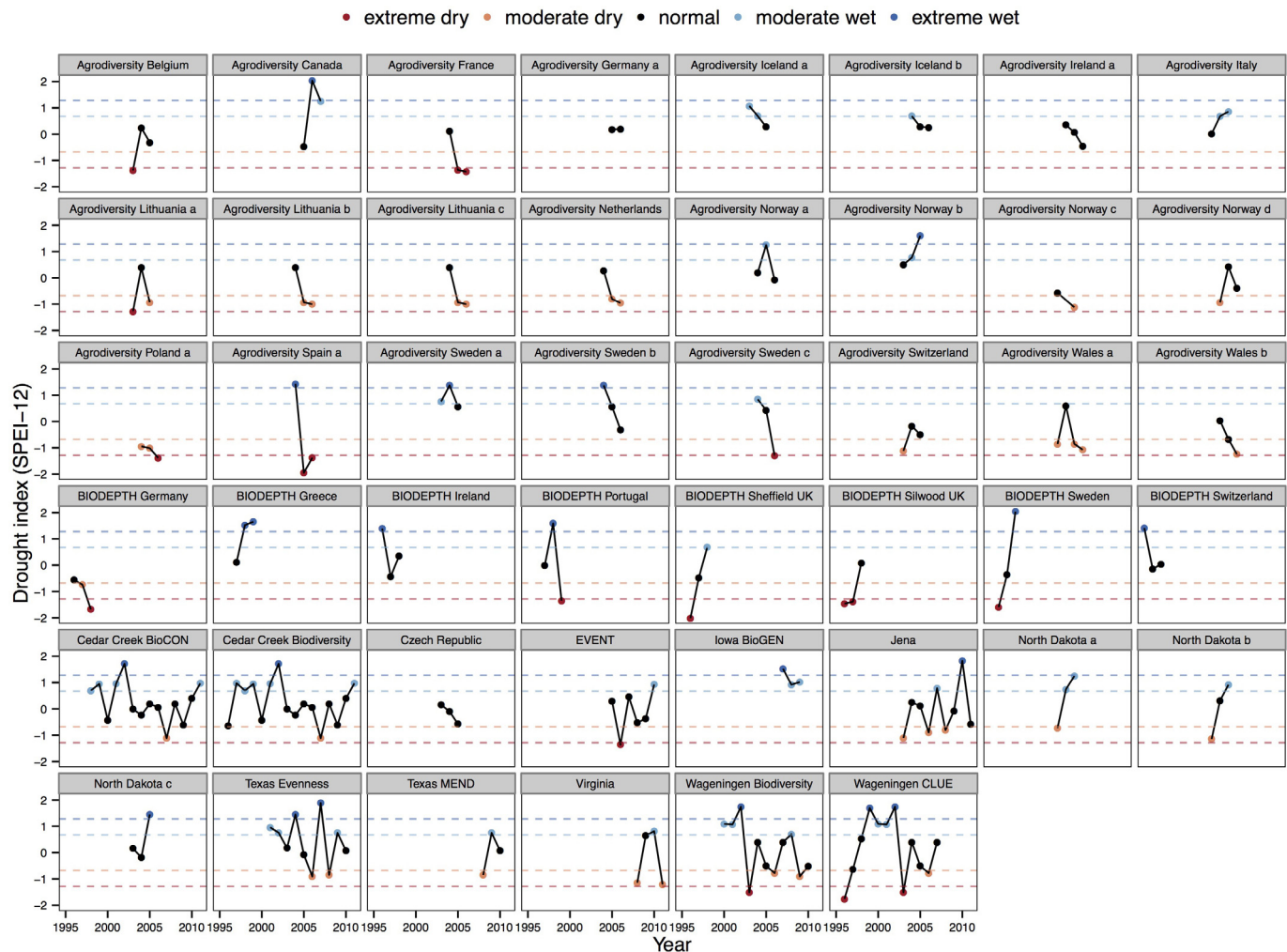
Extended Data Figure 3 | Classification of extreme dry, moderate dry, normal, moderate wet, and extreme wet years for each year of the 46 experiments. The 12-month version of the SPEI is shown, where positive values indicate wetter than normal water balances (precipitation minus potential evapotranspiration) during the 12-month time interval preceding and including the month of peak biomass harvest. For example, if peak biomass was harvested in September, then SPEI-12 accounts for the water balance from the previous October to September. Drought index values are based on month-by-month variations in climate over the past century (January 1901 to December 2011), based on monthly means of measurements made at more than 4,000 weather stations worldwide, and provided on 0.5 degree × 0.5 degree grids globally. Dashed lines show cutoffs for 1 in 4 (±0.67) or 1 in 10 (±1.28) year events. Seven experiments that included only normal years (Agrodiversity Germany a, Agrodiversity Ireland a, Czech Republic) or that did not include any normal years (Agrodiversity Poland a, Agrodiversity Spain a, Iowa BioGEN, North Dakota a) were excluded from subsequent analyses because it was not possible to compare perturbed with normal productivity levels for these studies.
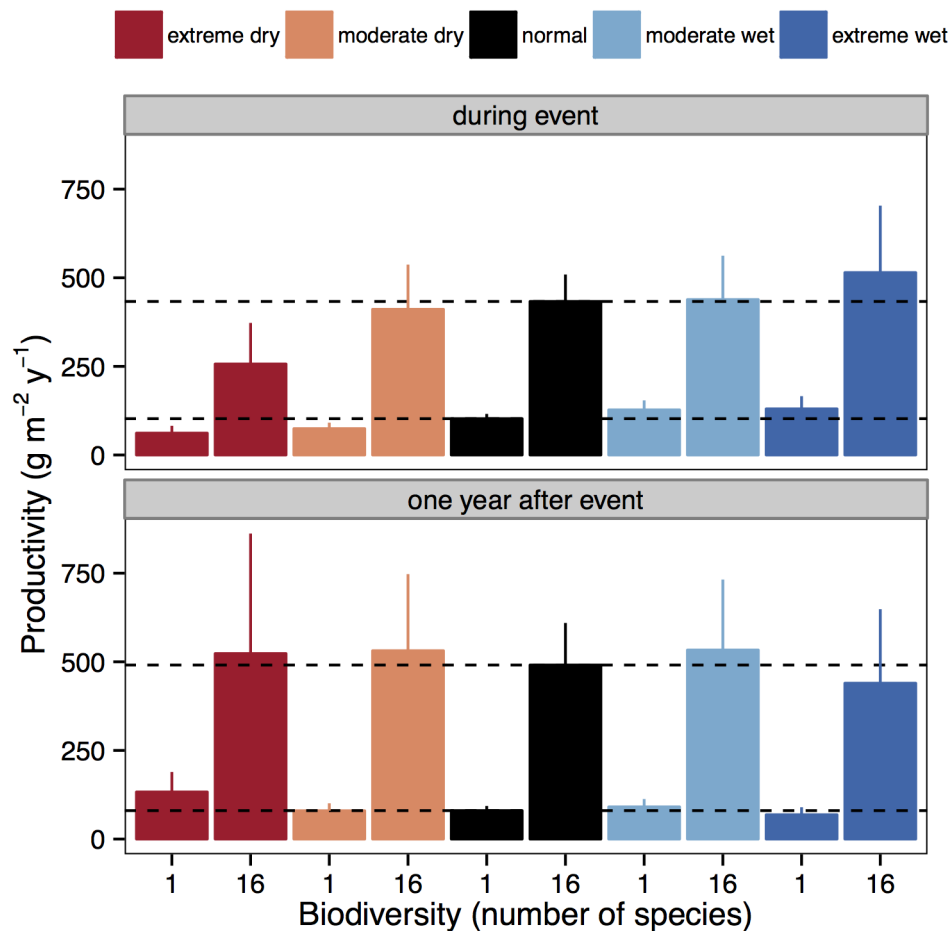
Extended Data Figure 4 | Productivity during and after both climate events and normal years for monocultures and mixtures of 16 species. Values shown are predicted means and 95% confidence intervals from the mixed-effects model. Productivity tends to be decreased during dry events and increased during wet events. This trend is reversed during the year after climate events. This pattern of overshooting normal levels of productivity during recovery 1 year after climate events is consistent with damped oscillations, rather than monotonic recovery (Extended Data Fig. 1). Relatively high productivity after extreme droughts could be due to increased nutrient availability and/or decreased abundance of herbivores as a result of reduced plant productivity during the drought. This might be especially true for low-diversity communities, which have the lowest productivity during drought, possibly explaining why biodiversity increases resilience after extremely dry years (Fig. 1c). Similarly, relatively low productivity after extremely wet years might be due to decreased nutrient availability and/or increased abundance of enemies as a result of increased plant productivity during the wet event. This might be especially true for high-diversity communities, which have the highest productivity during wet years, possibly explaining why biodiversity decreases resilience after extremely wet years (Fig. 1c). Dashed horizontal lines show normal productivity levels.
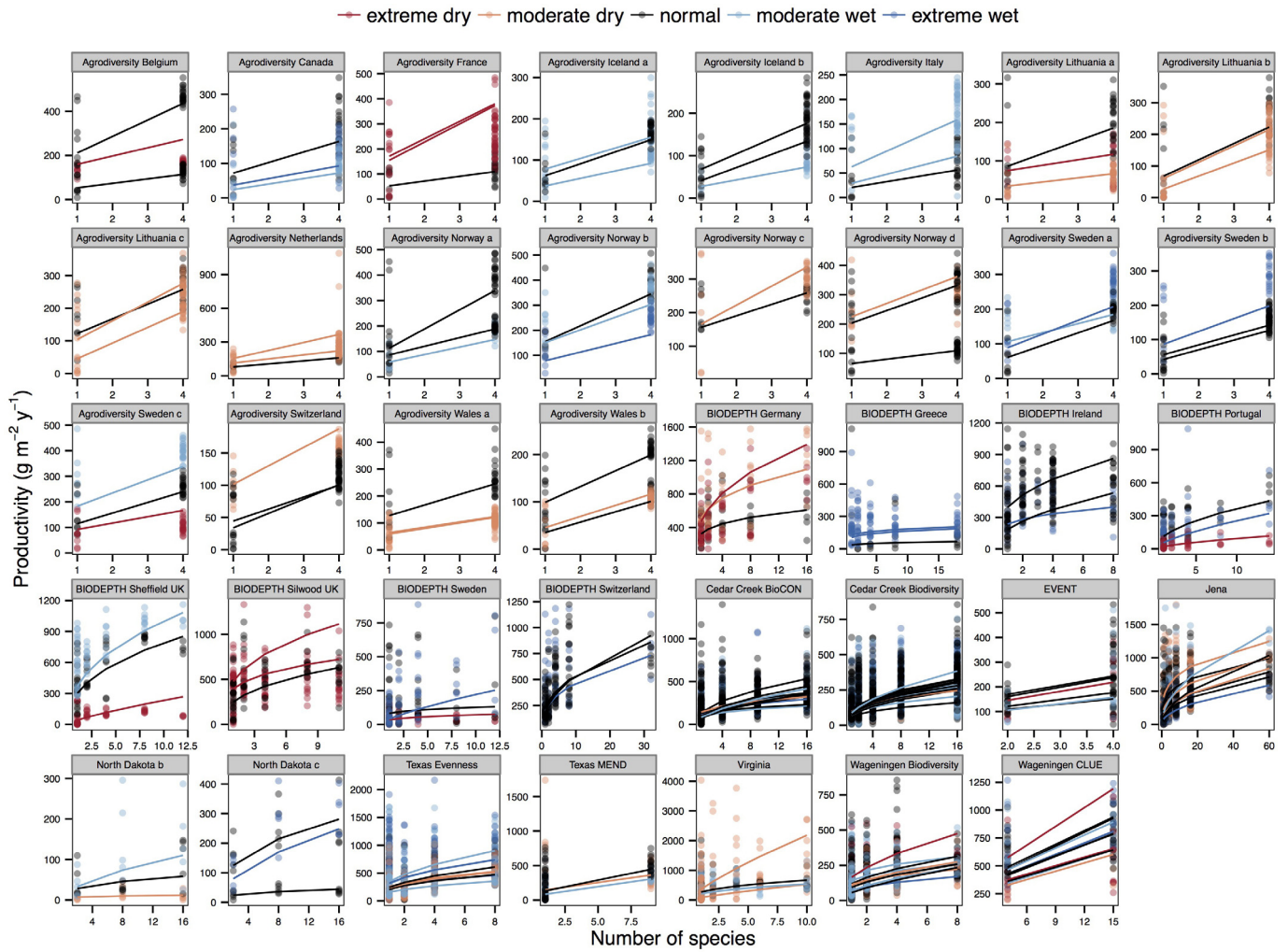
Extended Data Figure 5 | Biodiversity–productivity relationships for each year of each study, including normal years and climate events. Points are plot-level values and lines are mixed-model fits (Fig. 3).
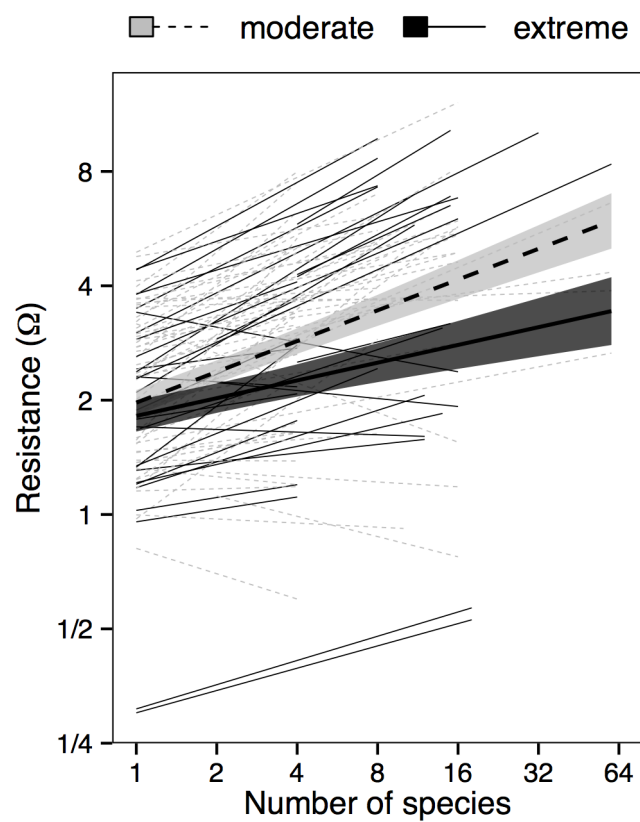
Extended Data Figure 6 | A marginally significant interaction between biodiversity and intensity (moderate or extreme). Table 1 indicates that productivity was marginally more resistant to moderate than to extreme climate events, especially at high biodiversity. All other interactions were non-significant ($P > 0.10$). Axes are logarithmic.
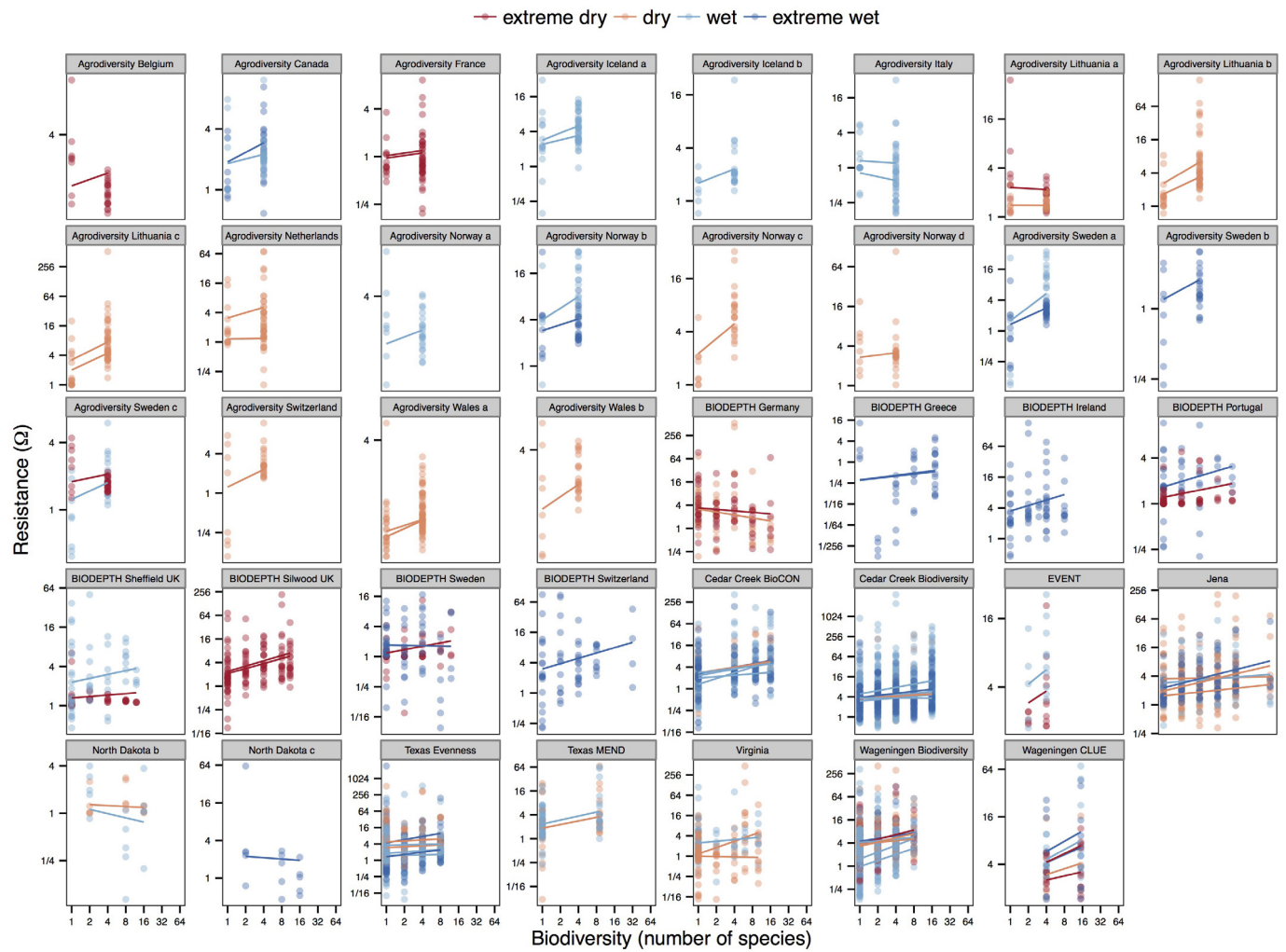
Extended Data Figure 7 | Biodiversity effects on the resistance of productivity to climate extremes. Shown for each study for which there were observations of productivity during both normal ($\overline{Y}_n$) and climate event (Y_e) years (Extended Data Fig. 3). Points are plot-level values and lines are mixed-model fits (Fig. 1b). Axes are logarithmic.
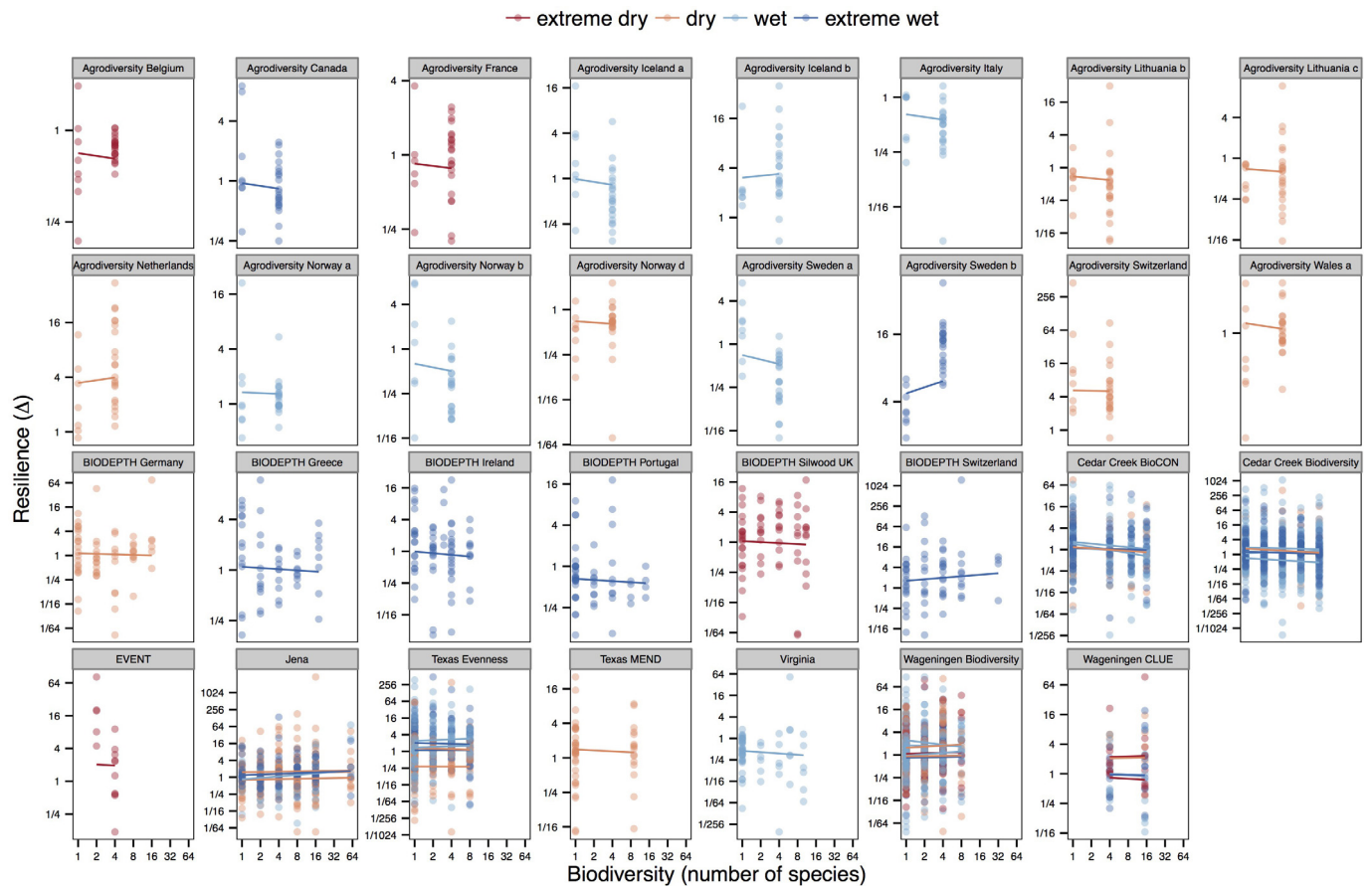
Extended Data Figure 8 | Biodiversity effects on the resilience of productivity to climate extremes. Shown for each study for which there were observations during normal ($\overline{Y_n}$), climate event (Y_e), and post-climate event (Y_{e+1}) years. Quantifying resilience requires more information (that is, Y_{e+1}) than quantifying resistance, thus we were unable to quantify resilience for eight of the studies shown in Extended Data Fig. 7. Specifically, we were unable to quantify resilience for studies where the only climate event occurred during the last year of the study (Extended Data Fig. 3) because in this case Y_{e+1} is unknown, and for studies where the only normal year was also the only post-event year (Extended Data Fig. 3) because in this case $\overline{Y_n} = Y_{e+1}$ and resilience is undefined. Points are plot-level values and lines are mixed-model fits (Fig. 1c). Axes are logarithmic.
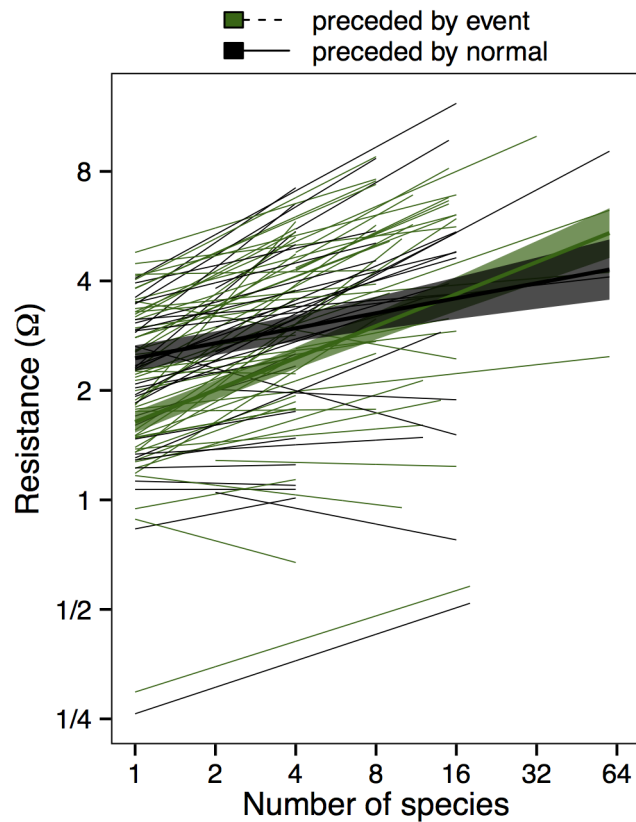
Extended Data Figure 9 | Biodiversity effects on the resistance of productivity to climate events that were preceded either by a climate event (green lines) or by a normal year (black lines). The significant interaction shown here indicates that biodiversity increased resistance more during climate events preceded by years with climate events than during climate events preceded by normal years ($F_{1,64.8} = 7.21$, $P < 0.01$). Axes are logarithmic. The sequence of climate events at each site is shown in Extended Data Fig. 3.

Extended Data Table 1 | Study details

Study	Years	# of years	Month of peak biomass harvest	# of plots	Levels of species richness
Agrodiversity Belgium	2003-2005	3	11	30	1,4
Agrodiversity Canada	2005-2007	3	8	30	1,4
Agrodiversity France	2004-2006	3	10	30	1,4
Agrodiversity Germany a	2005-2006	2	10	30	1,4
Agrodiversity Iceland a	2003-2005	3	8	30	1,4
Agrodiversity Iceland b	2004-2006	3	8	30	1,4
Agrodiversity Ireland a	2004-2006	3	11	29	1,4
Agrodiversity Italy	2003-2005	3	12	30	1,4
Agrodiversity Lithuania a	2003-2005	3	10	30	1,4
Agrodiversity Lithuania b	2004-2006	3	10	30	1,4
Agrodiversity Lithuania c	2004-2006	3	10	30	1,4
Agrodiversity Netherlands	2004-2006	3	10	30	1,4
Agrodiversity Norway a	2004-2006	3	8	30	1,4
Agrodiversity Norway b	2003-2005	3	9	30	1,4
Agrodiversity Norway c	2003-2005	2	10	30	1,4
Agrodiversity Norway d	2004-2006	3	8	30	1,4
Agrodiversity Poland a	2004-2006	3	10	30	1,4
Agrodiversity Spain a	2004-2006	3	7	30	1,4
Agrodiversity Sweden a	2003-2005	3	9	30	1,4
Agrodiversity Sweden b	2004-2006	3	9	30	1,4
Agrodiversity Sweden c	2004-2006	3	9	30	1,4
Agrodiversity Switzerland	2003-2005	3	10	30	1,4
Agrodiversity Wales a	2003-2006	4	10	30	1,4
Agrodiversity Wales b	2004-2006	3	11	30	1,4
BIODEPTH Germany	1996-1998	3	8	60	1,2,4,8,16
BIODEPTH Greece	1997-1999	3	5	52	1,2,4,8,18
BIODEPTH Ireland	1996-1998	3	8	70	1,2,3,4,8
BIODEPTH Portugal	1997-1999	3	5	56	1,2,4,8,14
BIODEPTH Sheffield UK	1996-1998	3	9	54	1,2,4,8,12
BIODEPTH Silwood UK	1996-1998	3	9	66	1,2,4,8,11
BIODEPTH Sweden	1996-1998	3	8	58	1,2,4,8,12
BIODEPTH Switzerland	1995-1997	3	8	64	1,2,4,8,32
Cedar Creek BioCON	1998-2011	14	8	74	1,4,9,16
Cedar Creek Biodiversity	1996-2011	16	8	168	1,2,4,8,16
Czech Republic	2003-2005	3	6	96	1,3,6,12
EVENT	2005-2010	6	9	15	2,4
Iowa BioGEN	2007-2009	3	8	64	1,4
Jena	2003-2011	9	9	82	1,2,4,8,16,60
North Dakota a	2003-2005	3	8	15	2,8,16
North Dakota b	2003-2005	3	8	15	2,8,16
North Dakota c	2003-2005	3	8	15	2,8,16
Texas Evenness	2001-2010	10	10	75	1,2,4,8
Texas MEND	2008-2010	3	10	52	1,9
Virginia	2008-2011	4	8	64	1,2,4,6,10
Wageningen Biodiversity	2000-2010	11	8	102	1,2,4,8
Wageningen CLUE	1996-2007	12	8	10	4,15

LETTER

Hedgehog actively maintains adult lung quiescence and regulates repair and regeneration

Tien Peng[1]†, David B. Frank[2], Rachel S. Kadzik[3], Michael P. Morley[1,4,5], Komal S. Rathi[1,4,5], Tao Wang[5], Su Zhou[5], Lan Cheng[5], Min Min Lu[5] & Edward E. Morrisey[1,3,4,5,6]

Postnatal tissue quiescence is thought to be a default state in the absence of a proliferative stimulus such as injury. Although previous studies have demonstrated that certain embryonic developmental programs are reactivated aberrantly in adult organs to drive repair and regeneration[1–3], it is not well understood how quiescence is maintained in organs such as the lung, which displays a remarkably low level of cellular turnover[4,5]. Here we demonstrate that quiescence in the adult lung is an actively maintained state and is regulated by hedgehog signalling. Epithelial-specific deletion of sonic hedgehog (Shh) during postnatal homeostasis in the murine lung results in a proliferative expansion of the adjacent lung mesenchyme. Hedgehog signalling is initially downregulated during the acute phase of epithelial injury as the mesenchyme proliferates in response, but returns to baseline during injury resolution as quiescence is restored. Activation of hedgehog during acute epithelial injury attenuates the proliferative expansion of the lung mesenchyme, whereas inactivation of hedgehog signalling prevents the restoration of quiescence during injury resolution. Finally, we show that hedgehog also regulates epithelial quiescence and regeneration in response to injury via a mesenchymal feedback mechanism. These results demonstrate that epithelial–mesenchymal interactions coordinated by hedgehog actively maintain postnatal tissue homeostasis, and deregulation of hedgehog during injury leads to aberrant repair and regeneration in the lung.

The Hedgehog (Hh) pathway coordinates tissue–tissue interactions in multiple organs during embryonic development through paracrine activation of smoothened (Smo)-mediated downstream signalling events[6,7]. We have previously demonstrated that Shh expressed by nascent lung endoderm progenitors coordinates cardiopulmonary mesoderm progenitor differentiation into various cardiac and lung mesenchymal cell lineages[8]. To determine whether Hh signalling continues to be active in the postnatal adult lung, we used the Shh^{creGFP} reporter[9] and our data show that Shh is expressed in the adult lung epithelium predominantly in the Scgb1a1$^+$ club epithelial cells in the proximal airway (Fig. 1a), with scattered expression in ciliated epithelium (Extended Data Fig. 1a) and the Sftpc$^+$ alveolar type II epithelial cells (Fig. 1b). The downstream transcriptional effector and target of hedgehog Gli1 (ref. 10), is expressed predominantly in mesenchymal cells adjacent to the proximal airway and pulmonary artery (Fig. 1c), with scattered expression in the alveolar interstitium as previously reported (Fig. 1d)[11]. Lineage tracing in the adult lung with $Gli1^{creERT2}$:$R26R^{mTmG}$ animals[12] showed that Gli1$^+$ Hh-responsive cells express several mesenchymal markers including Pdgfra, Pdgfrb, vimentin, S100A4, and Col1a1 (Fig. 1e–h, Extended Data Fig. 1b, c). Gli1$^+$ Hh-responsive mesenchymal cells do not contribute markedly to the smooth muscle lineage under homeostatic conditions, with the exception of rare venous smooth muscle within the proximal pulmonary venous myocardium (Extended Data Fig. 1d–i) and
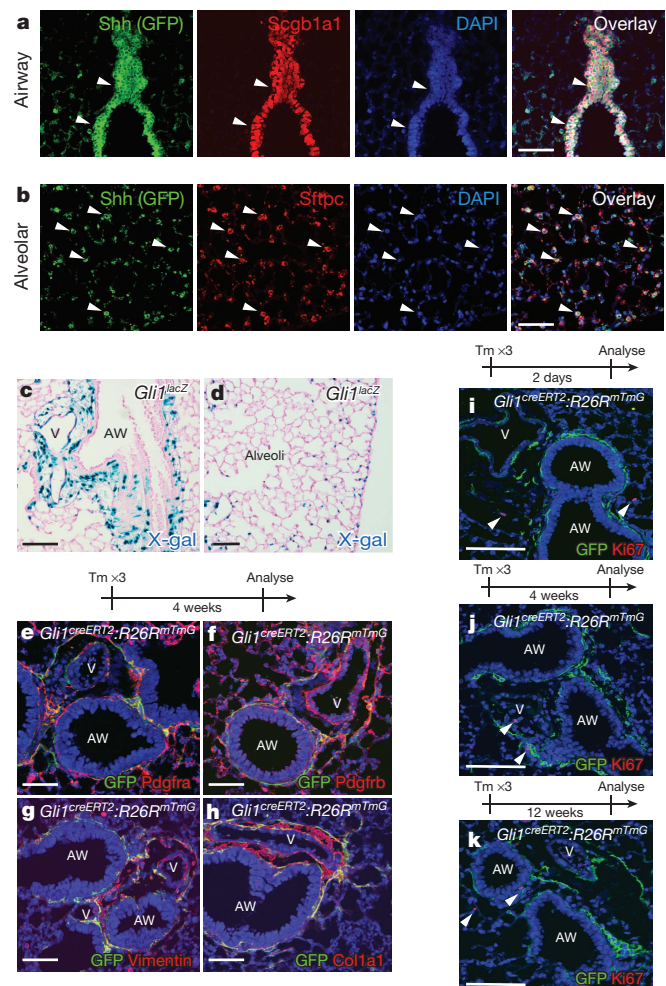

Figure 1 | The lung epithelium signals to the adjacent mesenchyme via paracrine Hh signalling during normal homeostasis. a, b, The Shh ligand is expressed in airway epithelium marked by Scgb1a1 (a), with scattered expression in the Sftpc$^+$ alveolar epithelium (b). c, d, The $Gli1^{LacZ}$ reporter is expressed in the mesenchyme adjacent to the airway and pulmonary artery (c) with scattered activation in the alveolar interstitium (d). e–h, Lineage traced Gli1$^+$ cells express Pdgfra$^+$ (e), Pdgfrb$^+$ (f), vimentin (g), and Col1a1 (h). i–k, Lineage-traced Gli1$^+$ cells do not expand in the adult lung after chase periods of 2 days (i), 4 weeks (j) and 12 weeks (k), with negligible expression of the cell cycle marker Ki67 (i–k, arrowheads). AW, airway; V, blood vessel, Tm, tamoxifen. Scale bars, 100 μm. Images representative of 3 animals with 5 sections examined per animal.

[1]Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [2]Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [3]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [4]Penn Center for Pulmonary Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [5]Penn Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [6]Penn Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. †Present address: University of California, San Francisco, 513 Parnassus Avenue, HSE Building, Room 1312, Box 0130, San Francisco, California 94143, USA (T.P.).
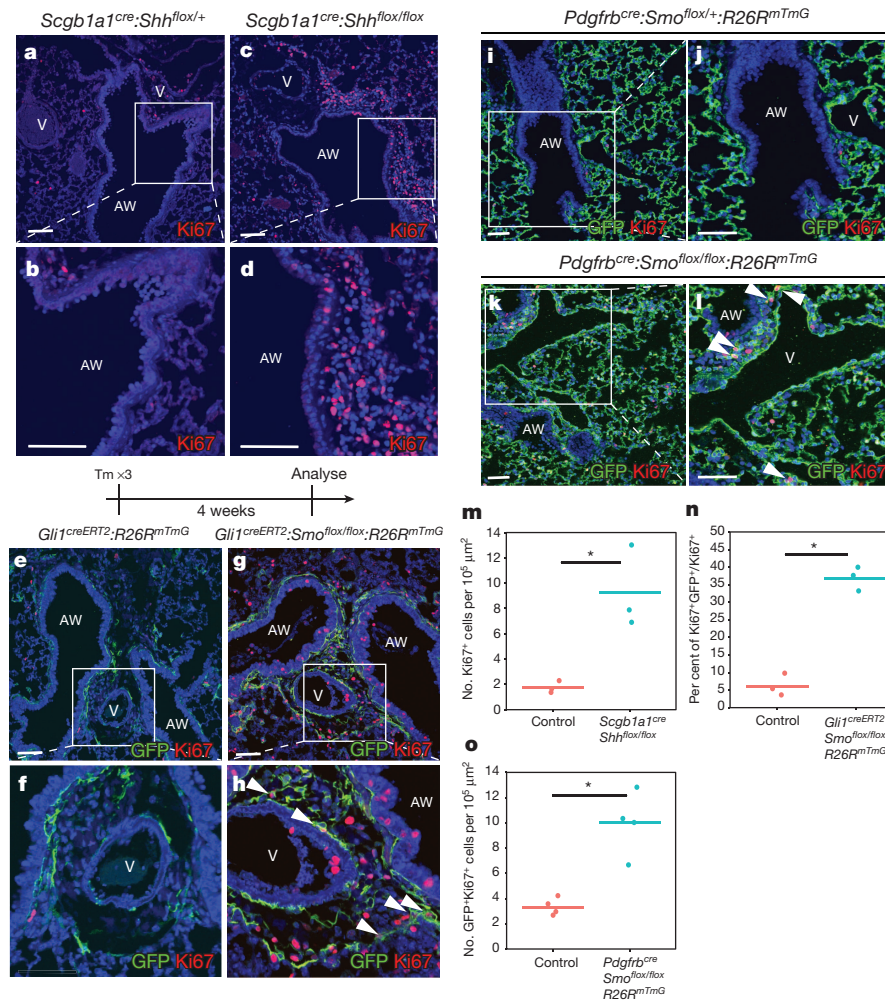
Figure 2 | Postnatal activation of Hh signalling is required to maintain lung mesenchymal quiescence. a–d, m, Deletion of Shh from airway epithelium increases proliferation in the mesenchyme surrounding the airway. **e–h, n**, Deletion of Smo within Gli1⁺ cells causes proliferation and mesenchymal expansion as noted by increased Ki67 expression. **i–l, o**, Deletion of Smo within Pdgfrb⁺ mesenchyme shows increased mesenchymal proliferation in the adult lung. AW, airway; V, blood vessel; CKO, conditional knockout; blue represents DAPI counterstaining. Scale bars, 100 μm; *P < 0.05. Data represent n = 3 animals per group with 5 sections analysed per animal. *In vitro* lung mesenchyme studies represent technical triplicates, with BrdU assay representative of three separate experiments. One-sided *t*-test used to determine statistical significance with centre value representing the mean and error bar representing s.d.

myofibroblasts during fibrotic injury (Extended Data Fig. 1j). Gli1⁺ cells do not contribute to cells of the haematopoietic lineage in the lung (Extended Data Fig. 1k). Gli1⁺ cells remain quiescent up to 12 weeks after lineage labelling, with little to no notable expansion or Ki67 labelling (Fig. 1i–k).

We deleted Shh using the *Scgb1a1^cre* driver, which is active in the airway epithelium, to define the importance of Hh signalling in the postnatal lung (Extended Data Fig. 2a, b)[13]. Examination of *Scgb1a1^cre:Shh^flox/flox* adult lungs reveals mesenchymal expansion and increased mesenchymal cell proliferation surrounding the airway epithelium (Fig. 2a–d, m and Extended Data Fig. 2c–h). Thus, epithelial-specific loss of Shh in the postnatal lung is sufficient to induce cellular proliferation in the adjacent mesenchyme.

To address the cell-autonomous role of Hh signalling in adult lung mesenchyme, we deleted Smo within Gli1⁺ Hh-responsive cells in the adult lung and followed their proliferative response. Four weeks after Smo deletion, lineage traced Gli1⁺ mesenchymal cells expanded relative to controls and exhibited increased cell proliferation (Fig. 2e–h, n). We also deleted Smo using the mesenchyme-specific *Pdgfrb^cre* driver[14], and *Pdgfrb^cre:Smo^flox/flox:R26R^mTmG* adult mutants exhibit increased cell proliferation and expansion of the Pdgfrb-derived population surrounding the airways and in the alveolar interstitium (Fig. 2i–l, o and

Extended Data Fig. 3a–i). Adult *Pdgfrb^cre:Smo^flox/flox:R26R^mTmG* mutants older than 6 months exhibit elevated pulmonary arterial pressures, indicating that loss of Hh signalling at the bronchovascular interface causes pulmonary hypertension (Extended Data Fig. 3j–l).

We then assessed the transcriptome of isolated adult lung mesenchymal cells expressing the activated SmoM2 mutant form of smoothened, resulting in increased Gli1 expression (Extended Data Fig. 4a–i)[15]. Unbiased gene ontology (GO) analysis showed highest enrichment in the subset of genes involved in "mitotic nuclear division", with most of these transcripts downregulated in Hh-activated fibroblasts (Extended Data Fig. 4j, Supplementary Tables 1 and 2), suggesting that Hh activation attenuates cell cycle progression in the adult lung mesenchyme.

Previous studies have demonstrated that Pdgfr signalling promotes postnatal mesenchymal proliferation[16,17], and that Pdgfr isoforms are expressed in the adult lung mesenchyme (Fig. 1, Extended Data Fig. 4c, d, j). Therefore, we assessed the interaction between Hh and Pdgf signalling using a gain-of-function mutant of Pdgfrb (*Pdgfrb^(S)K*, hereafter referred to as *Pdgfrb^GOF*)[17]. Activation of Pdgfrb within Hh-responsive Gli1⁺ lung mesenchymal cells results in their proliferative expansion (Extended Data Fig. 4m–p, s). However, concurrent expression of SmoM2 attenuates the Pdgfrb-induced expansion of
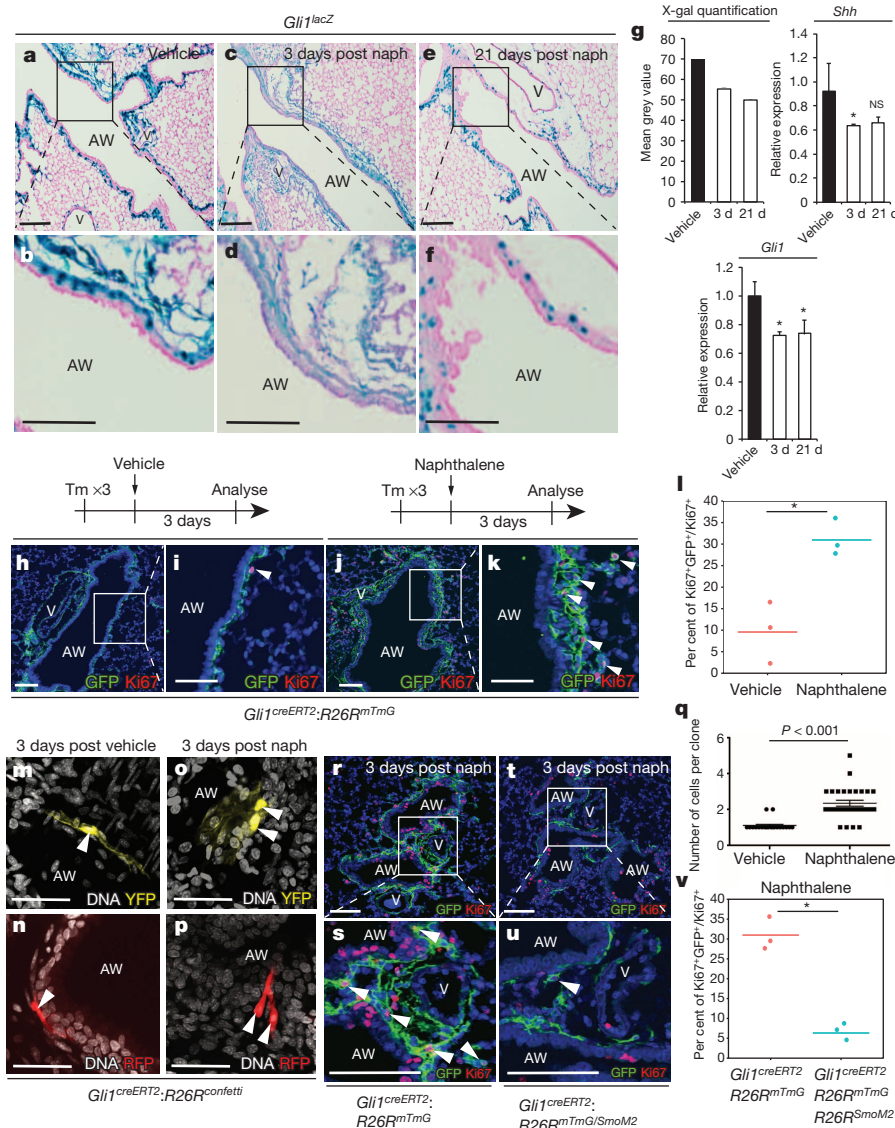
Figure 3 | Hh signalling modulates the acute mesenchymal response to epithelial injury. a–f, Naphthalene injury downregulates Gli1[lacZ] expression (n = 2 animals per group) and **g**, Shh and Gli1 expression noted by quantitative PCR (n = 3 animals per group). NS, not significant. **h–l**, Gli1[+] lung cells undergo proliferative expansion shortly after naphthalene injury as measured by Ki67[+] expression. **m–q**, Clonal analysis of Gli1[+] lung cells at single cell resolution demonstrates clonal expansion three days after injury.

r–v, Conditional activation of Smo (SmoM2) within lineage-traced Gli1[+] lung cells attenuates the proliferative expansion that follows epithelial injury with naphthalene. AW, airway; V, blood vessel. Scale bars, 100 μm; *P < 0.05. Blue staining represents DAPI (**h–k, q–t**) except in confetti experiments where white represents TO-PRO-3 DNA counterstaining (**m–p**). Data represent n = 3 animals per group with 5 sections analysed per animal. Clonal analysis represents >50 clones analysed in 4 animals. Error bars, mean ± s.d.

Gli1[+] Hh-responsive mesenchyme (Fig. 2q–s). Activation of Hh signalling in isolated lung mesenchymal cells in vitro (derived from UBC[creERT2]:R26R[SmoM2] animals) attenuates the proliferation induced by exogenous Pdgf-BB (platelet-derived growth factor-BB) (Extended Data Fig. 4l).

Next, we assessed the expression of Hh signalling components during airway epithelial injury with naphthalene[18]. Acute naphthalene injury caused a reduction in Hh activation as assessed by decreased Gli1[LacZ] reporter activity in the mesenchyme surrounding the airway, reduced expression of Shh and Gli1 transcripts, and decreased expression of GFP in the Shh[creGFP] reporter (Fig. 3a–g, Extended Data Fig. 5). Chronic repetitive bleomycin caused a similar reduction in Hh activation following injury (Extended Data Figs 5 and 6). Thus, Hh signalling is downregulated in response to epithelial injury in the lung, and is not upregulated as has been previously reported[19,20]. Of note, these results correlate with the loss of Shh-expressing epithelium after injury.

To assess the behaviour of Gli1[+] lung cells after epithelial injury, we exposed Gli1[creERT2]:R26R[mTmG] adult animals to tamoxifen followed by a one-week washout period before inducing lung epithelial injury with naphthalene. Hh-activated Gli1[+] lung cells rapidly undergo proliferative expansion after naphthalene injury (Fig. 3h–l). Utilizing the Gli1[creERT2]:R26R[confetti] mice for stochastic multicolour clonal analysis, we demonstrate that individual Gli1[+] cells clonally expand after naphthalene injury (Fig. 3m–q). Reconstitution of Hh activation with SmoM2 during acute epithelial injury attenuates the normal expansion of mesenchyme following injury (Fig. 3r–v). In the bleomycin injury model, Hh signalling is also downregulated within Gli1[+] mesenchymal cells after injury, which is similarly attenuated by the expression of activated SmoM2 (Extended Data Fig. 6).

Despite an initial reduction in Hh activation during naphthalene injury, Shh and Gli1 expression return to homeostatic levels three months following injury (Fig. 4a–e) as the Shh-expressing bronchial epithelium is reconstituted (Extended Data Fig. 7a–d). Mesenchymal
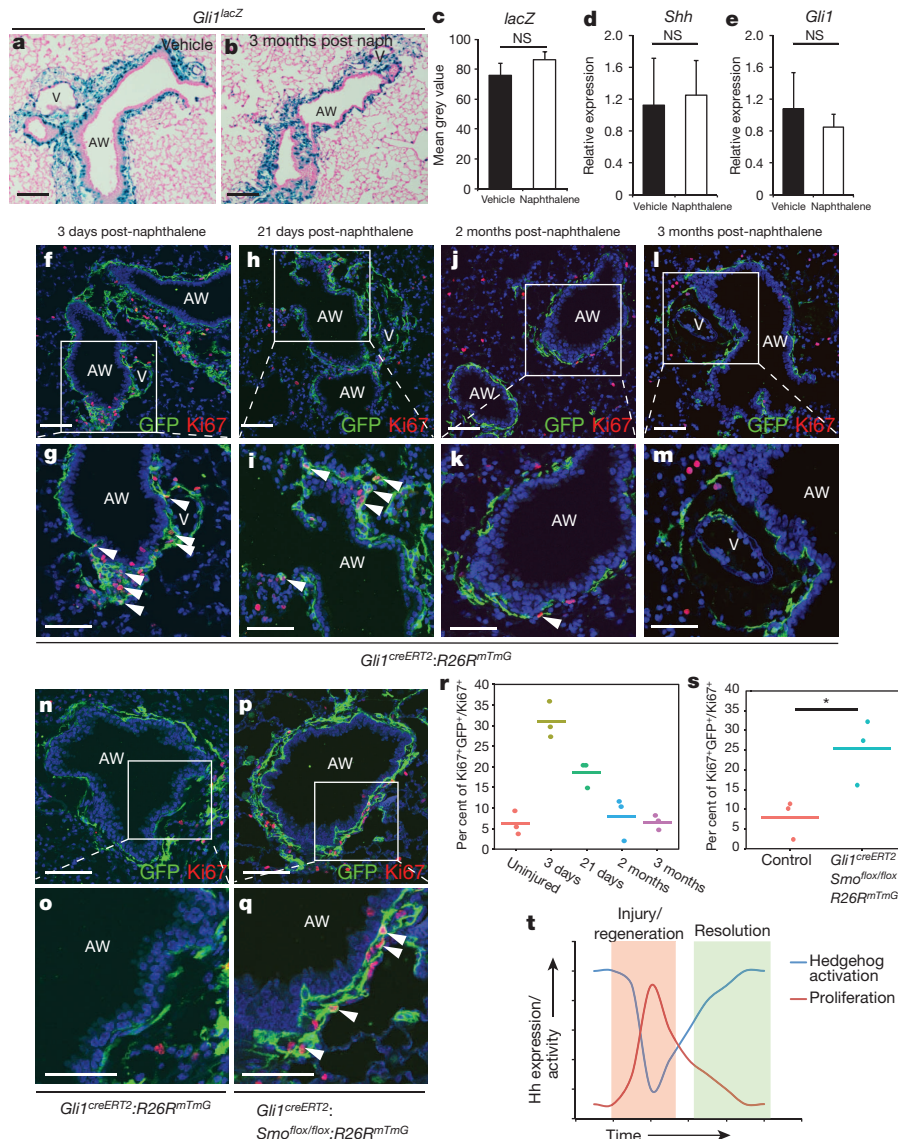
Figure 4 | Hh signalling modulates restoration of quiescence during injury resolution in the lung. a–c, Hh activation returns to homeostatic levels three months after naphthalene injury in *Gli1^{lacZ}* lungs (n = 3 animals per group, 5 sections analysed per animal) and **d, e,** *Shh* and *Gli1* expression returns to normal as noted by qPCR (n = 3 animals per group). **f–m, r,** Gli1$^+$ lung cells undergo proliferative expansion shortly after naphthalene injury but return to quiescence by 2–3 months after injury. **n–q, s,** Conditional deletion of Smo in

lineage-traced Gli1$^+$ lung cells prevents the restoration of mesenchymal quiescence. **t,** Relationship between Hh expression and activation and lung mesenchymal proliferation as plotted over time during injury and resolution. AW, airway; V, blood vessel. Scale bars, 100 μm; *P < 0.05. Data represent n = 3 animals per group with 5 sections analysed per animal unless otherwise noted. Error bars, mean ± s.d.

quiescence is also gradually restored after 2–3 months (Fig. 4f–m, r). Deletion of Smo within Gli1$^+$ Hh-responsive cells prevented the restoration of mesenchymal quiescence as these Gli1$^+$ mesenchymal cells surrounding the airways continue to proliferate 2 months after naphthalene injury (Fig. 4n–q, s). These data show that Hh activation is dynamically regulated after epithelial injury and is inversely correlated with mesenchymal proliferation as injury repair and regeneration progresses (Fig. 4t).

Bronchial Scgb1a1$^+$ secretory cells have tremendous proliferative capacity to regenerate damaged epithelial airways after injury[21,22] (Extended Data Fig. 7a–d). Therefore, we assessed whether Hh activation in the mesenchyme alters secretory epithelial proliferation and regeneration in conditional Hh loss and gain of function mutants. *Scgb1a1^{cre}:Shh^{flox/flox}* and *Pdgfrb^{cre}:Smo^{flox/flox}:R26R^{mTmG}* mutants demonstrate a significant increase in bronchial epithelial proliferation (Fig. 5a–f), while the *Gli1^{creERT2}:Smo^{flox/flox}:R26R^{mTmG}* mutants show a trend towards increased epithelial proliferation during normal

homeostasis (Fig. 5g–i). To determine whether mesenchymal activation of Hh signalling affects epithelial proliferation and regeneration after injury, we activated and deleted Smo within Gli1$^+$ mesenchyme during naphthalene-induced epithelial injury. Activation of Hh results in a marked loss of Scgb1a1$^+$ secretory epithelium 2 months after naphthalene injury relative to multi-ciliated epithelium (TubbIV$^+$) (Fig. 5k, m), which does not undergo cellular turnover with naphthalene injury[23]. In contrast, inactivation of Hh signalling promotes excessive Scgb1a1$^+$ club cell regeneration, leading to bronchial hyperplasia (Fig. 5l, m). Next, we generated lung organoids from Scgb1a1-derived epithelium and cultured it in the presence or absence of isolated lung mesenchyme (Extended Data Fig. 7e–n). Organoids co-cultured with mesenchyme predominantly formed colonies expressing markers of the secretory lineage with a small fraction generating alveolar epithelial cells, while those without mesenchyme failed to form colonies (Extended Data Fig. 7e–k). Activation of SmoM2 in the co-cultured lung mesenchyme reduced the number and
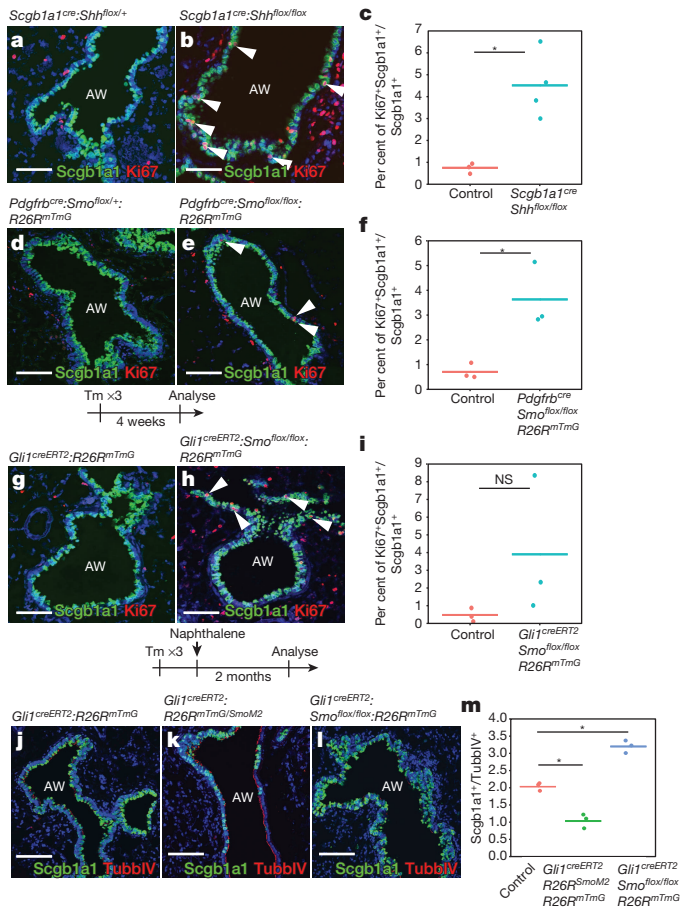
Figure 5 | Hh signalling regulates epithelial quiescence via mesenchymal feedback. a–f, Deletion of Shh from the proximal secretory epithelium and deletion of Smo within cells derived from Pdgfrb[+] mesenchyme increase proliferation of Scgb1a1[+] club cells during homeostasis. **g–i,** Inducible deletion of Smo within Gli1[+] lung mesenchymal cells results in a non-significant trend towards increased Scgb1a1[+] club cell proliferation. **j–m,** Activation of SmoM2 in Gli1[+] mesenchyme results in impaired regeneration of Scgb1a1[+] cells after naphthalene injury, whereas deletion of Smo in the mesenchyme induces excessive expansion of Scgb1a1[+] cells resulting in bronchial hyperplasia. AW, airway. Scale bars, 100 μm; *$P < 0.05$. Data represent $n = 3$ animals per group with 5 sections analysed per animal. Error bars, mean ± s.d.

size of the epithelial colonies (Extended Data Fig. 7l–n). These data show that Hh promotes epithelial quiescence via a mesenchymal feedback mechanism, possibly by downregulating stromal factors necessary for epithelial proliferation.

In this study, we have demonstrated that the lung epithelium actively maintains mesenchymal quiescence through paracrine Hh signalling, which also regulates a feedback loop to maintain epithelial quiescence. This finding stands in contrast to the known role of Shh in promoting cell proliferation during tissue development as well as its role in promoting tumorigenesis in adults. While previous reports have suggested that Hh signalling is pro-mitogenic in the adult lung[11,19,20], our study is the first report to utilize multiple genetic models to assess Hh function in the adult lung *in vivo* and *in vitro*. Our data indicate that certain signalling pathways such as Hh maintain a balance between proliferation and quiescence during lung homeostasis and regeneration. Our studies reveal that disruption of this balance upon injury can lead to changes in expansion of the mesenchyme, which may disrupt epithelial regeneration after injury or in disease (Extended Data Fig. 8).

1. Beers, M. F. & Morrisey, E. E. The three R's of lung health and repair, remodeling, and regeneration. *J. Clin. Invest.* **121,** 2065–2073 (2011).
2. Herriges, M. & Morrisey, E. E. Lung development: orchestrating the generation of a complex organ. *Development* **141,** 502–513 (2014).
3. Hogan, B. L. *et al.* Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15,** 123–138 (2014).
4. Blenkinsopp, W. K. Proliferation of respiratory tract epithelium in the rat. *Exp. Cell Res.* **46,** 144–154 (1967).
5. Breuer, R., Zajicek, G., Christensen, T. G., Lucey, E. C. & Snider, G. L. Cell kinetics of normal adult hamster bronchial epithelium in the steady state. *Am. J. Respir. Cell Mol. Biol.* **2,** 51–58 (1990).
6. McMahon, A. P., Ingham, P. W. & Tabin, C. J. Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53,** 1–114 (2003).
7. Lum, L. & Beachy, P. A. The Hedgehog response network: sensors, switches, and routers. *Science* **304,** 1755–1759 (2004).
8. Peng, T. *et al.* Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. *Nature* **500,** 589–592 (2013).
9. Harfe, B. D. *et al.* Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118,** 517–528 (2004).
10. Bai, C. B., Auerbach, W., Lee, J. S., Stephen, D. & Joyner, A. L. Gli2, but not Gli1, is required for initial Shh signaling and ectopic activation of the Shh pathway. *Development* **129,** 4753–4761 (2002).
11. Liu, L. *et al.* Hedgehog signaling in neonatal and adult lung. *Am. J. Respir. Cell Mol. Biol.* **48,** 703–710 (2013).
12. Ahn, S. & Joyner, A. L. *In vivo* analysis of quiescent adult neural stem cells responding to Sonic hedgehog. *Nature* **437,** 894–897 (2005).
13. Li, H. *et al.* Cre-mediated recombination in mouse Clara cells. *Genesis* **46,** 300–307 (2008).
14. Foo, S. S. *et al.* Ephrin-B2 controls cell motility and adhesion during blood-vessel-wall assembly. *Cell* **124,** 161–173 (2006).
15. Jeong, J., Mao, J., Tenzen, T., Kottmann, A. H. & McMahon, A. P. Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.* **18,** 937–951 (2004).
16. Olson, L. E. & Soriano, P. Increased PDGFRα activation disrupts connective tissue development and drives systemic fibrosis. *Dev. Cell* **16,** 303–313 (2009).
17. Olson, L. E. & Soriano, P. PDGFRβ signaling regulates mural cell plasticity and inhibits fat development. *Dev. Cell* **20,** 815–826 (2011).
18. Buckpitt, A. R., Bahnson, L. S. & Franklin, R. B. Comparison of the arachidonic acid and NADPH-dependent microsomal metabolism of naphthalene and 2-methylnaphthalene and the effect of indomethacin on the bronchiolar necrosis. *Biochem. Pharmacol.* **35,** 645–650 (1986).
19. Bolaños, A. L. *et al.* Role of Sonic Hedgehog in idiopathic pulmonary fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **303,** L978–L990 (2012).
20. Watkins, D. N. *et al.* Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* **422,** 313–317 (2003).
21. Zemke, A. C. *et al.* β-Catenin is not necessary for maintenance or repair of the bronchiolar epithelium. *Am. J. Respir. Cell Mol. Biol.* **41,** 535–543 (2009).
22. Wang, Y. *et al.* Development and regeneration of Sox2+ endoderm progenitors are regulated by a Hdac1/2-Bmp4/Rb1 regulatory pathway. *Dev. Cell* **24,** 345–358 (2013).
23. Rawlins, E. L., Ostrowski, L. E., Randell, S. H. & Hogan, B. L. Lung development and repair: contribution of the ciliated lineage. *Proc. Natl Acad. Sci. USA* **104,** 410–417 (2007).

Author Contributions T.P. and E.E.M. designed the overall experimental strategy. T.P. and R.S.K. performed lineage tracing and animal injury experiments. T.P. and D.B.F. performed *in vitro* BrdU and organoid experiments. T.W. performed right heart catheterization on the animals. S.Z., L.C. and M.M.L. performed histology and immunohistochemistry. T.P., D.B.F., K.S.R., M.P.M. and E.E.M. analysed the data. T.P. and E.E.M. wrote and edited the manuscript with input from all authors.

Author Information The Gene Expression Omnibus accession number for the microarray data produced in these studies is GSE68201. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.E.M. (emorrise@mail.med.upenn.edu) or T.P. (tien.peng@ucsf.edu).

METHODS

No statistical methods were used to predetermine sample size.

Animals. Generation and genotyping of the Shh^{creGFP} (ref. 9), $Gli1^{LacZ}$ (ref. 10), $Gli1^{creERT2}$ (ref. 12), $Scgb1a1^{cre}$ (ref. 13), $Pdgfrb^{cre}$ (ref. 14), $R26R^{confetti}$ (ref. 24), $Smo^{flox/flox}$ (ref. 25), $Shh^{flox/flox}$ (ref. 26), $R26R^{mTmG}$ (ref. 27), $R26R^{SmoM2}$ (ref. 15), $Pdgfrb^{(S)K}$ (ref. 17), $UBC^{CreERT2}$ (ref. 28) lines have been previously described. The animals were housed and treated in accordance with the IACUC protocol approved at the University of Pennsylvania. Animals between the ages of 8–12 weeks old were used for the experiments with balance of gender between groups. Tamoxifen (Sigma) was dissolved in corn oil and administered intraperitoneally at 200 mg kg^{-1} per day × 3 days for lineage tracing studies, with the exception of clonal analysis studies with the $R26R^{confetti}$ reporter, where only one dose of tamoxifen was given at 200 mg kg^{-1}.

Histological analysis. Mouse lungs were inflated and fixed in 2% paraformaldehyde, dehydrated in a series of increasing ethanol concentration washes, embedded in paraffin and sectioned. Antibodies used were anti-sm22α (goat anti-SM22α 1:200 Abcam), GFP (goat anti-GFP 1:100 Abcam, rabbit anti-GFP 1:100 Molecular Probe), Scgb1a1 (goat anti-Scgb1a1 1:20 Santa Cruz), SPC (rabbit-anti SPC 1:500 Chemicon), Pdgfra (rabbit anti-Pdgfra 1:50 Cell Signaling), Pdgfrb (rabbit anti-Pdgfrb 1:100 Cell Signaling), vimentin (rabbit anti-vimentin 1:100 Santa Crux), collagen type 1 (rabbit anti-Col1 1:500 Abcam), Ki67 (rabbit anti-Ki67 1:50 Abcam), PCNA (mouse anti-PCNA 1:50 Biocare), PO4-Histone H3 (mouse anti-PO4-Histone H3 1:200 Cell Signaling), TubbIV (mouse anti-TubbIV 1:20 Biogenex), S100A4 (rabbit anti-S100A4 1:200 Abcam). LacZ staining of lungs was performed as previously described[8]. The slide was imaged on a Zeiss LSM 710 confocal microscope and analysed in ImageJ software.

Animal injury experiments. For acute naphthalene injury, mice were given 300 mg kg^{-1} of naphthalene (Sigma) dissolved in corn oil via intraperitoneal injection. For chronic bleomycin injury, mice were given 50 U kg^{-1} of pharmaceutical grade bleomycin (Hospira) dissolved in PBS via intraperitoneal injection twice a week for four weeks.

Measurement of pulmonary artery pressure. Following anaesthesia with Avertin, the trachea was cannulated, and mice were ventilated using a MiniVent Type 845 (Harvard Apparatus). The chest cavity was opened to expose the heart, and a Micro-Tip Catheter Transducer SPR-1000 (Millar Instruments) was inserted into the right ventricle. Systolic right ventricle pressure was measured as a surrogate for systolic pulmonary artery pressure, recorded on a PowerLab 4/30 instrument (ADInstruments), and analysed using Chart 5 Pro software (ADInstruments). Pressure measurements associated with heart rates outside the range of 300–500 beats per minute were excluded from analysis. For each mouse, 2–4 measurements were analysed, each corresponding to the average of 10–20 individual data points. The experimenter was blinded to the mouse genotype and three mice of each genotype were examined.

Clonal analysis of $Gli1^{creERT2}$:$R26R^{confetti}$ lungs. For clonal analysis of $Gli1^{creERT2}$:$R26R^{confetti}$ mice, lungs were inflated and fixed in 2% PFA overnight, washed with cold PBS four times, and then cleared using the Sc*a*le reagent as reported[29]. Clarified lung specimen was then counterstained with TO-PRO3 (Life Technologies) for nuclear counterstaining and dissected into slices ~1 mm thick and mounted on a Fastwell with coverslip and sealed. Sections were imaged on a Zeiss LSM 710 confocal microscope and analysed in ImageJ software. Thick sections were randomly sampled for single-coloured clones with identical-colour labelled cells within 50 μm of each other considered as derived from the same clone. Colour and spatially-segregated clones of 1–5 cells were identified and plotted in a box plot according to experimental conditions (vehicle versus naphthalene).

Cell counting and image analysis. Sections included in cell count analysis were acquired using confocal microscopy. At least four animals per genotype were used. Cell counts were performed on ImageJ using the "Cell Counter" plug-in and the performer was blinded to the specimen genotype and condition. Results were averaged between each specimen and standard deviations were calculated per genotype. One-tailed paired *t*-tests were used to determine the *P* value. Quantification of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)-positive or GFP$^+$ pixels in lung sections was performed using ImageJ. Lung sections were captured on a Nikon Eclipse light microscope under identical exposures and converted to monochrome 8-bit images, inverted, and the mean grey value was quantified over the X-gal- or GFP-positive area surrounding the airway and vasculature.

qPCR. Total RNA was isolated from whole lung or cultured primary lung fibroblasts using the RNeasy kit (Qiagen) and following the manufacturer's protocol. Complementary DNA was synthesized from total RNA using the SuperScript Strand Synthesis System (Invitrogen). Quantitative PCR was performed using the SYBR Green system (Applied Biosystems) with the following primers:
Shh F′ 5′-AAGTACGGCCATGCTGGCTCGC-3′
Shh R′ 5′-QCCACGGAGTTCTCTGCTTTCACAG-3′
Gli1 F′ 5′-GTGCACGTTTGAAGGCTGTC-3′

Gli1 R′ 5′-TAAAGGCCTTGCTGCAACCT-3′
GAPDH F′ 5′-CCCCAGCAAGGACACTGAGCAAGAG-3′
GAPDH R′ 5′-GGCCCCTCCTGTTATTATGGGGGGT-3′

GAPDH expression values were used to control for RNA quantity. Data are shown as the average of a minimum of three biological replicates for each genotype per condition ± s.d.

Isolation and culture of lung mesenchymal cells. Whole lung was dissected from C57BL6 male adult animals and tracheally perfused with a digestion cocktail of Collagenase Type I (450 U ml^{-1}, Gibco), elastase (4 U ml^{-1}, Worthington) and dispase (1:10 BD Bioscience) and removed from the chest. The lung was further diced with razor blades and the mixture incubated at 37 °C for 25 min and vortexed intermittently. The mixture was then washed with DMEM-F12 and incubated with 0.1% trypsin-EDTA for 20 min and vortexed intermittently. The mixture was passed through a 100-μm cell strainer and resuspended in RBC lysis buffer, before passing through a 40-μm cell strainer. The resuspended cells were cultured on gelatin-treated tissue culture plates with DMEM-F12 plus 10% fetal calf serum. Media was refreshed every other day and primary lung mesenchymal cells were maintained for no more than three passages.

Microarray. Primary lung mesenchymal cells were isolated from $UBC^{creERT2}$:$R26R^{SmoM2}$ adult mice and grown in DMEM-F12 plus 10% fetal calf serum. The cells were treated with vehicle or 1 μg ml^{-1} of 4-OH-tamoxifen in DMEM F12 without serum and total RNA was isolated after 48 h. Biotinylated cRNA probe libraries were generated from these RNA samples and assayed with the Affymetrix Mouse Gene 2.0ST genechip. Microarray data were analysed using the Oligo package available at the Bioconductor Website (http://www.bioconductor.org). The raw data were background-corrected by the robust multichip average (RMA) method and then normalized by an invariant set method. Genes with 80% of samples with an expression signal above the negative control probes were considered detectable or present. Differential gene expression analysis between control and mutant mice was analysed by the Limma package available at the Bioconductor Website. *P* values were adjusted for multiple comparison using a false discovery rate. GO enrichment analysis was performed using the Bioconductor package topGO. The Gene Expression Omnibus accession number for the microarray data produced in these studies is GSE68201.

Cell proliferation assay. Lung mesenchymal cells were isolated from $UBC^{creERT2}$:$R26R^{SmoM2}$ animals and plated at 1×10^4 cells per well in 96 well plates and grown for 3 days with vehicle or 1 μg ml^{-1} of 4-OH-tamoxifen until cells became confluent. Cells were then incubated in serum-free DMEM F12 for 24 h before Pdgf-BB (mouse, R&D) was added and cultured for another 24 h. BrdU was then added to the media after 24 h of Pdgf-BB incubation and BrdU incorporation was assayed after four hours according to manufacturer instructions (Cell Signaling Technology, BrdU cell proliferation assay kit).

Bronchial organoid formation assay. GFP$^+$ bronchial epithelium were FACS sorted from $Scgb1a1^{cre}$:$R26R^{mTmG}$ lungs and co-cultured with lung mesenchyme isolated from $UBC^{creERT2}$:$R26R^{SmoM2}$ animals (5×10^3 epithelial cells to 5×10^4 mesenchymal cells per well) in a modified MTEC media diluted 1:1 in growth factor reduced Matrigel (Corning). Modified MTEC culture media is comprised of small airway basal media (SABM) (Lonza) with selected components from SAGM bullet kit (Lonza) including insulin, transferrin, bovine pituitary extract, retinoic acid, and gentamicin/amphotericin B. Additional components include 25 ng ml^{-1} mEGF (Sigma), 0.1 μg ml^{-1} cholera toxin (Sigma), and 5% FBS (Life Technologies). Cell suspension-Matrigel mixture was placed in a transwell and incubated in growth media with 10 μM ROCK inhibitor (Sigma) in a 24-well plate with vehicle or 1 ug ml^{-1} 4-OH-tamoxifen for 48 h, after which the media was replenished every 48 h (lacking tamoxifen). Colonies were assayed after 14 days. Each experimental condition was performed in quadruplicates and counted blinded to the experimental condition. Colony forming efficiency = (number of GFP$^+$ colonies/number GFP$^+$ epithelial cells cultured per well) × 100. Areas of individual colonies were assayed on ImageJ and over 140 colonies were randomly sized per experimental condition.

24. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
25. Long, F., Zhang, X. M., Karp, S., Yang, Y. & McMahon, A. P. Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development* **128**, 5099–5108 (2001).
26. Lewis, P. M. et al. Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**, 599–612 (2001).
27. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
28. Ruzankina, Y. et al. Deletion of the developmentally essential gene ATR in adult mice leads to age-related phenotypes and stem cell loss. *Cell Stem Cell* **1**, 113–126 (2007).
29. Hama, H. et al. Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nature Neurosci.* **14**, 1481–1488 (2011).
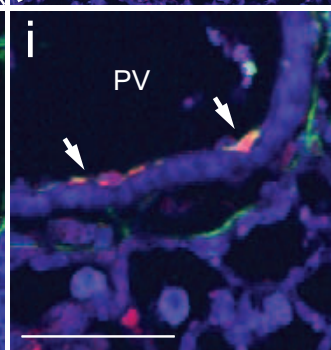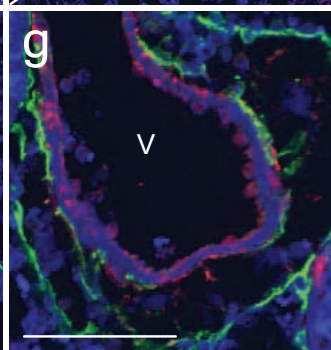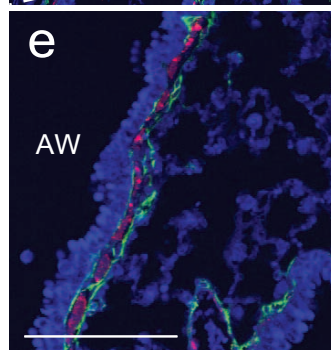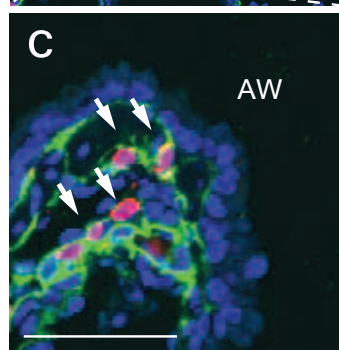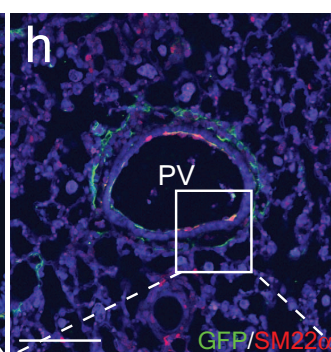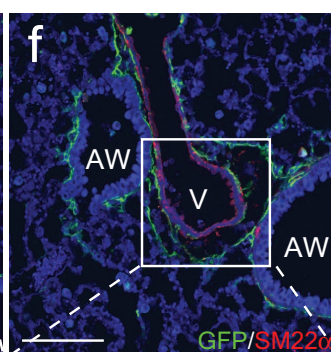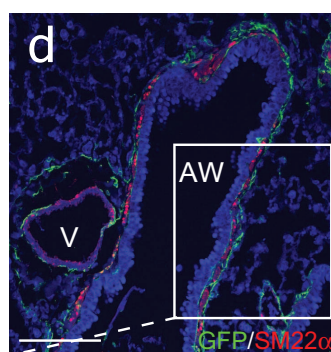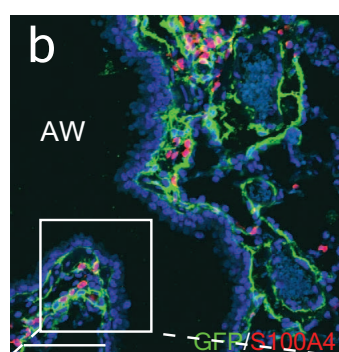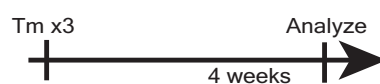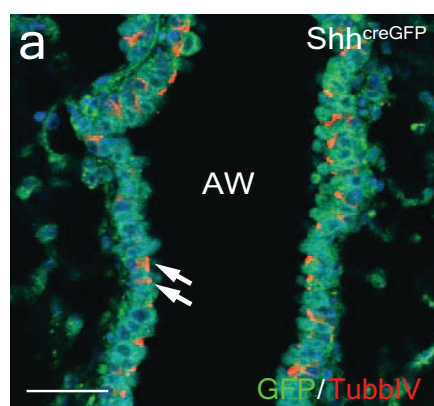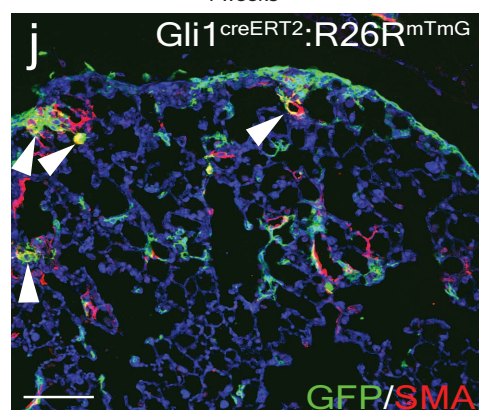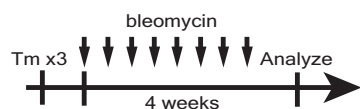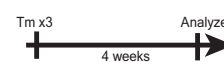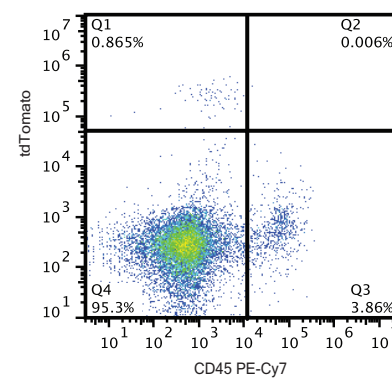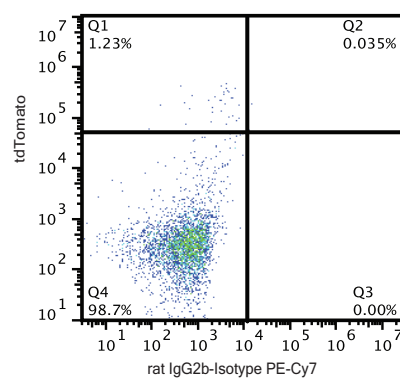
Extended Data Figure 1 | Characterization of Hh signalling in the lung. a, A small number of TubbIV$^+$ ciliated cells express GFP in the *Shh^{creGFP}* reporter. **b**, **c**, Gli1$^+$ Hh activated cells co-label with S100A4, a fibroblast marker. **d**–**g**, Lineage tracing 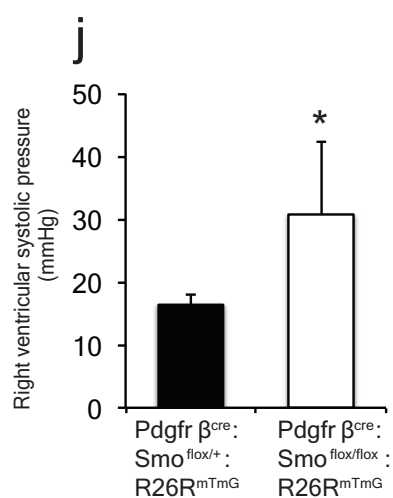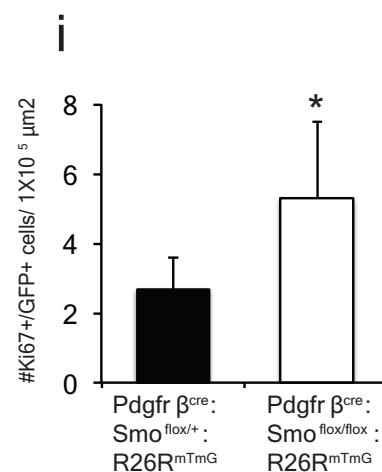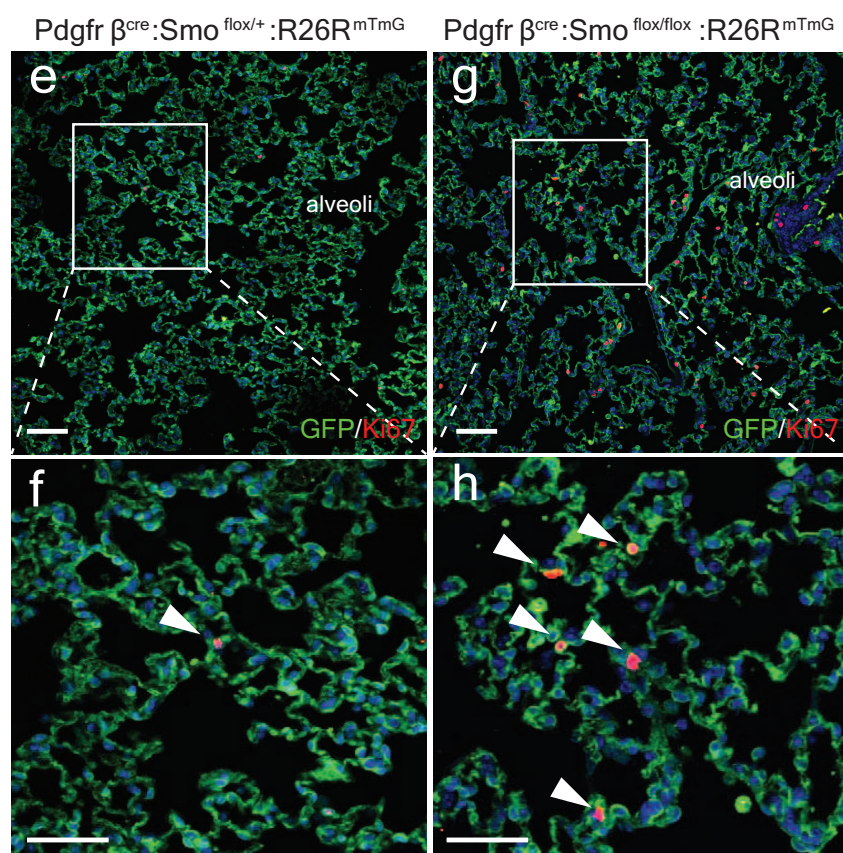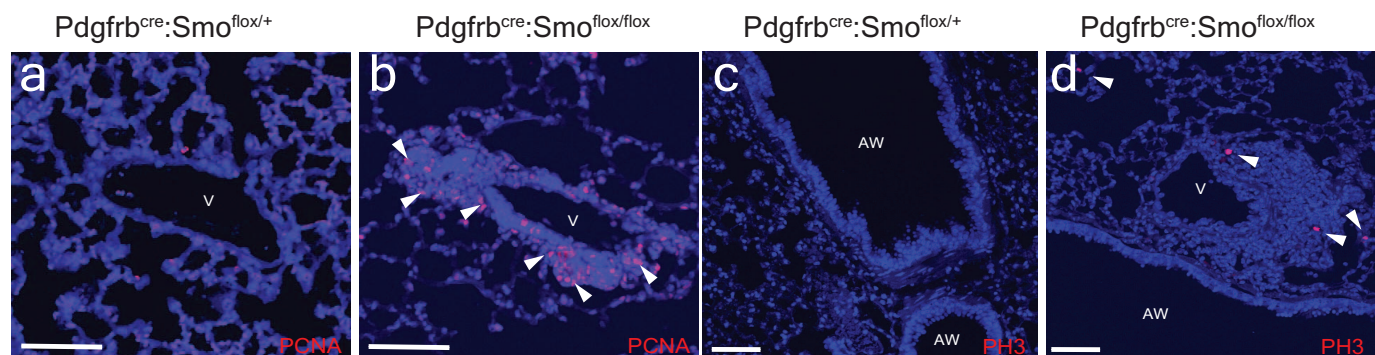of Gli1$^+$ Hh activated cells shows little to no co-localization of GFP$^+$ cells with the airway smooth muscle (**d**, **e**) or the vascular smooth muscle of the adjacent pulmonary artery (**f**, **g**). **h**, **i**, The rare exception occurs in the pulmonary vein where Gli1$^+$ Hh activated cells contribute to the venous smooth muscle that is surrounded by the venous myocardium of the proximal pulmonary vein. **j**, Gli1$^+$ cells also generate myofibroblasts after fibrotic injury such as that induced by bleomycin. **k**, Lung Gli1$^+$ cells do not contribute to cells of haematopoietic lineage as marked by CD45. AW, airway; V, blood vessel; PV, pulmonary vein. Scale bars, 100 μm. Images representative of 3 animals with 5 sections examined per animal.
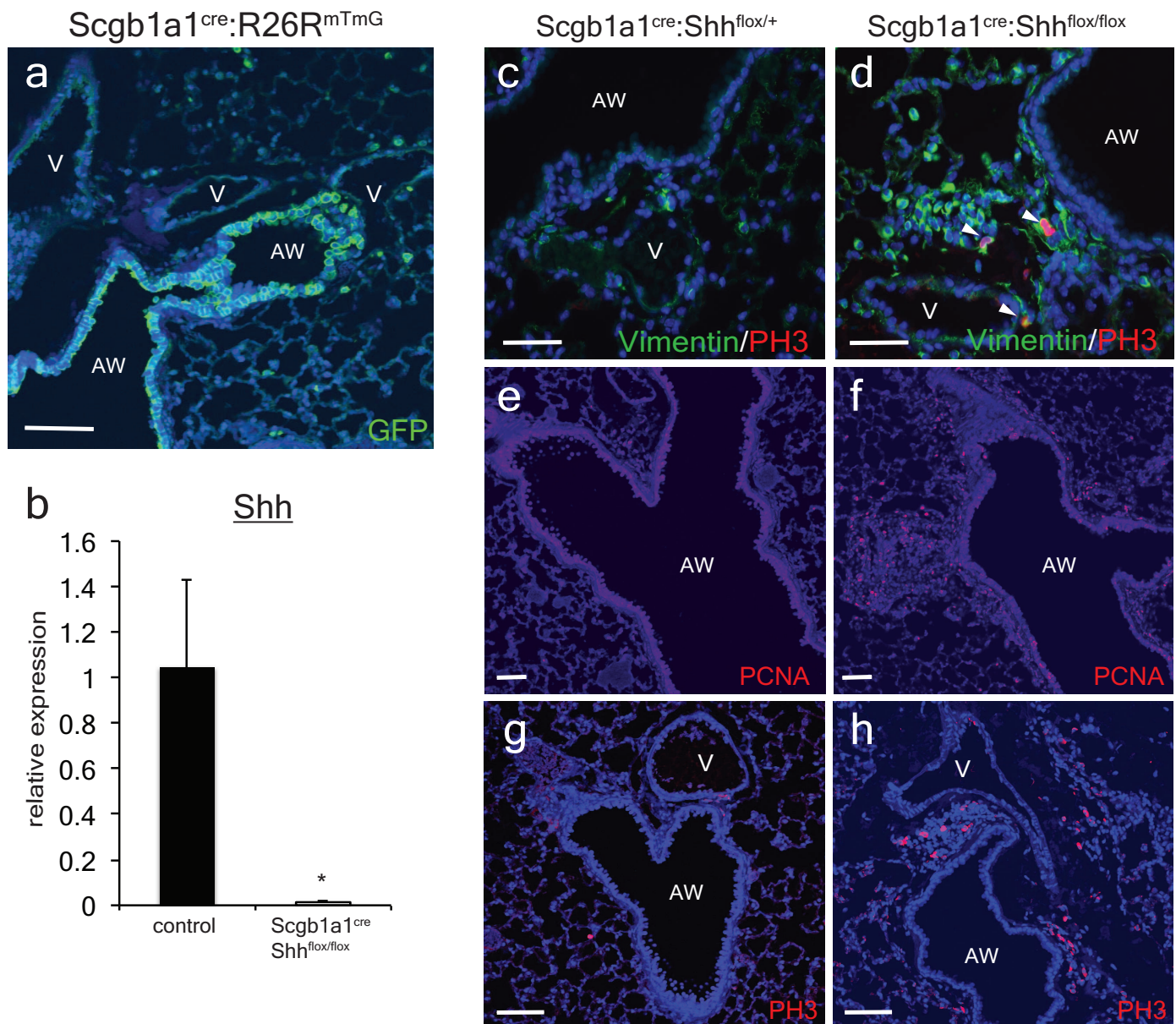
Extended Data Figure 2 | Conditional deletion of Shh from the adult airway epithelium increases proliferation in the adjacent mesenchyme. a, The *Scgb1a1*[cre] driver predominantly marks the airway epithelium in the adult lung when crossed with the *R26R*[mTmG] reporter. **b**, Whole-lung messenger RNA transcript analysis reveals efficient deletion of Shh transcripts in the *Scgb1a1*[cre]:*Shh*[flox/flox] animals compared to controls (*Shh*[flox/flox])

($n = 4$ animals). **c–h**, Deletion of Shh from the airway epithelium resulted in an increased expression of proliferative markers, PCNA and phospho-histone H3 (PH3) in the mesenchyme surrounding the airways in *Scgb1a1*[cre]:*Shh*[flox/flox] mutants (**d, f, h**, $n = 4$ animals) versus controls (*Scgb1a1*[Cre]:*Shh*[flox/+]) (**c, e, g**, $n = 3$ animals). AW, airway; V, blood vessel. Scale bars, 100 μm. *$P < 0.05$. Error bars, mean ± s.d.

Pdgfrb^cre:Smo^flox/+   Pdgfrb^cre:Smo^flox/flox   Pdgfrb^cre:Smo^flox/+   Pdgfrb^cre:Smo^flox/flox

Pdgfr β^cre:Smo^flox/+:R26R^mTmG   Pdgfr β^cre:Smo^flox/flox:R26R^mTmG

Pdgfr β^cre:Smo^flox/+:R26R^mTmG   Pdgfr β^cre:Smo^flox/flox:R26R^mTmG
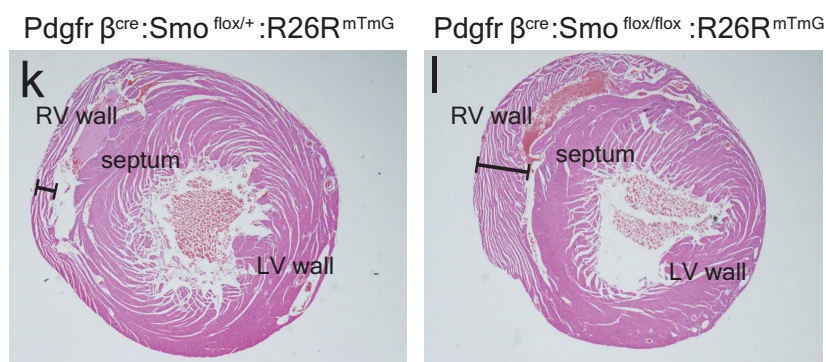
Extended Data Figure 3 | Conditional deletion of Smo from Pdgfrb-derived mesenchyme increases mesenchymal proliferation at the epithelial–mesenchymal interface and vascular remodelling. a–d, Deletion of Smo from Pdgfrb-derived mesenchyme resulted in increased expression of proliferative markers, PCNA and PH3 in the mesenchyme at the epithelial-mesenchymal interface of *Pdgfrb^cre^:Smo^flox/flox^* mutants (**b, d**) versus controls (*Pdgfrb ^Cre^:Smo^flox/+^*) (**a, c**). **e–i,** *Pdgfrb^cre^:Smo^flox/flox^:R26R^mTmG/+^* mutants exhibit increased Ki67$^+$ cells within lineage traced GFP$^+$ cells in the alveoli compared to controls (*n* = 4 animals per group). **j, k,** Aged *Pdgfrb^cre^:Smo^flox/flox^:R26R^mTmG/+^* mutants (>6 months old) spontaneously develop pulmonary hypertension with increased right ventricular systolic pressure (**j**, *n* = 3 animals per group) and right ventricle wall thickness (**k, l**). Scale bars, 100 μm. *$P < 0.05$. Error bars, mean ± s.d.
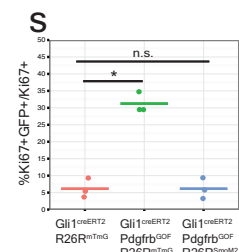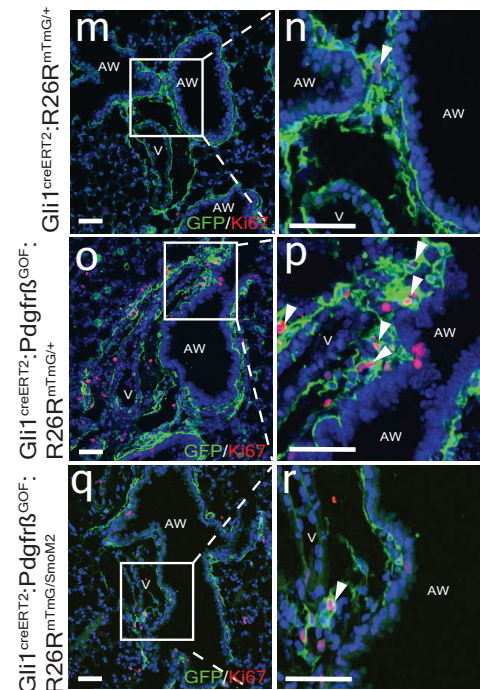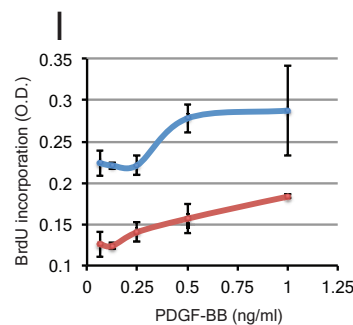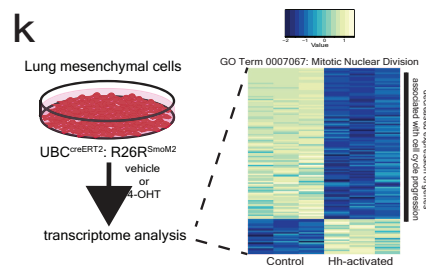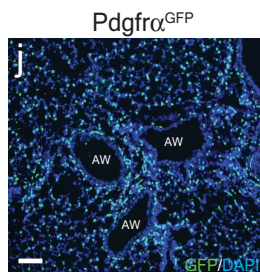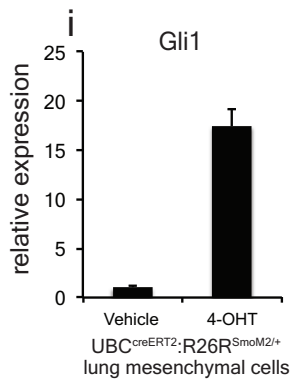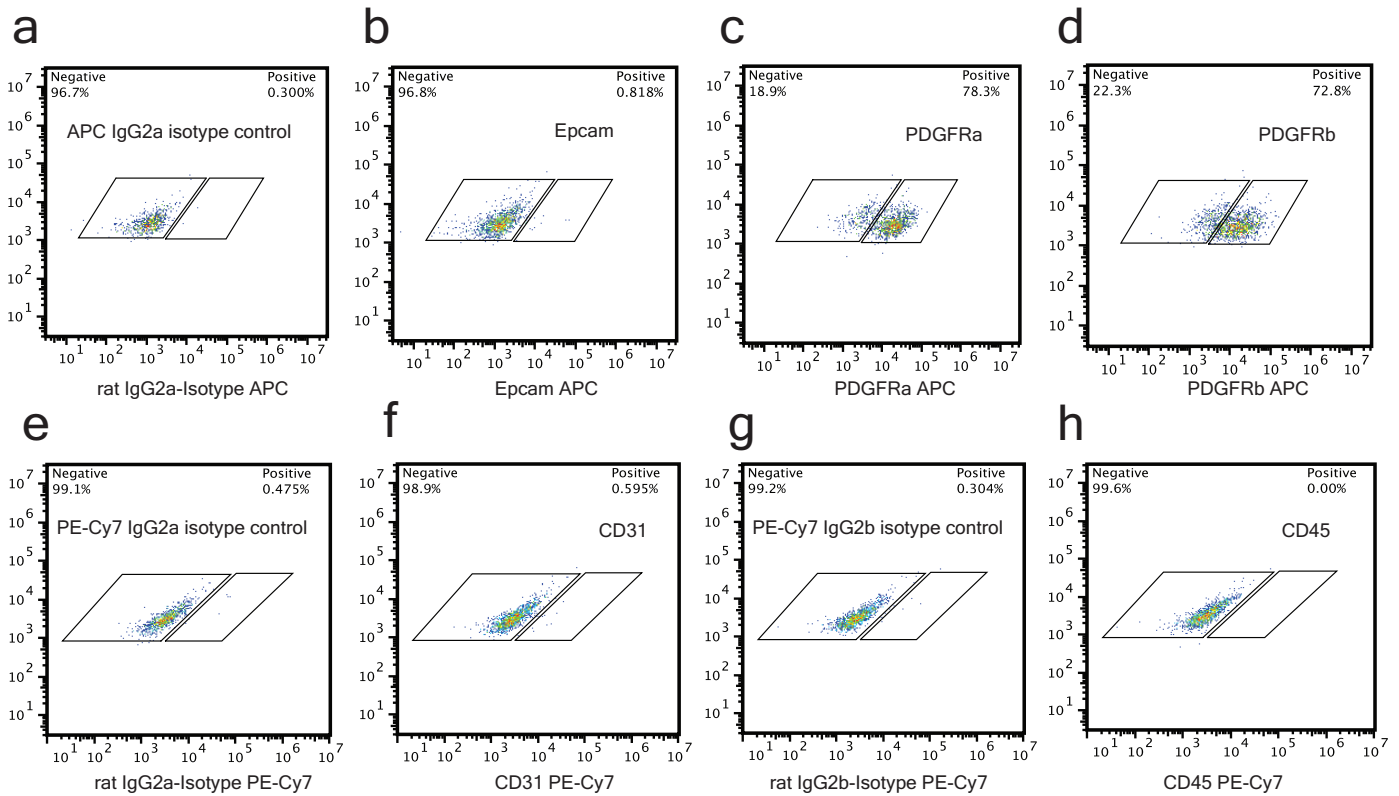
Extended Data Figure 4 | Characterization of isolated lung mesenchyme.
a–h, Isolated mesenchymal cells *in vitro* predominantly express Pdgfra and
Pdgfrb (**c**, **d**, compared to isotype control **a**), but not epithelial marker, Epcam
(**b**, compared to isotype control **a**), endothelial marker, CD31 (**f**, compared
to isotype control **e**), nor haematopoietic marker, CD45 (**h**, compared to
isotype control **g**). **i**, Expression of the constitutively active form of Smo
(SmoM2) by 4-OH-tamoxifen induction significantly upregulates Gli1
expression in the isolated lung mesenchyme after 48 h. **j**, GFP staining of the
Pdgfra^{GFP} reporter demonstrates that Pdgfra$^+$ cells are expressed broadly in the
lung. **k**, Activation of Smo in isolated lung mesenchymal cells leads to
reduced expression of cell cycle progression genes. **l**, Hh activation of lung
mesenchyme with SmoM2 attenuated the proliferation induced by Pdgf-BB
ligand *in vitro* as assayed by BrdU incorporation. **m–p, s**, Expression of
activated Pdgfrb(*Pdgfrb*^{GOF}) within Gli1$^+$ cells resulted in their proliferative
expansion. **q–s**, Concurrent activation of SmoM2 attenuated the proliferative
expansion induced by *Pdgfrb*^{GOF} ($n = 3$ animals per group). Scale bars, 100 μm.
Error bars, mean ± s.d.
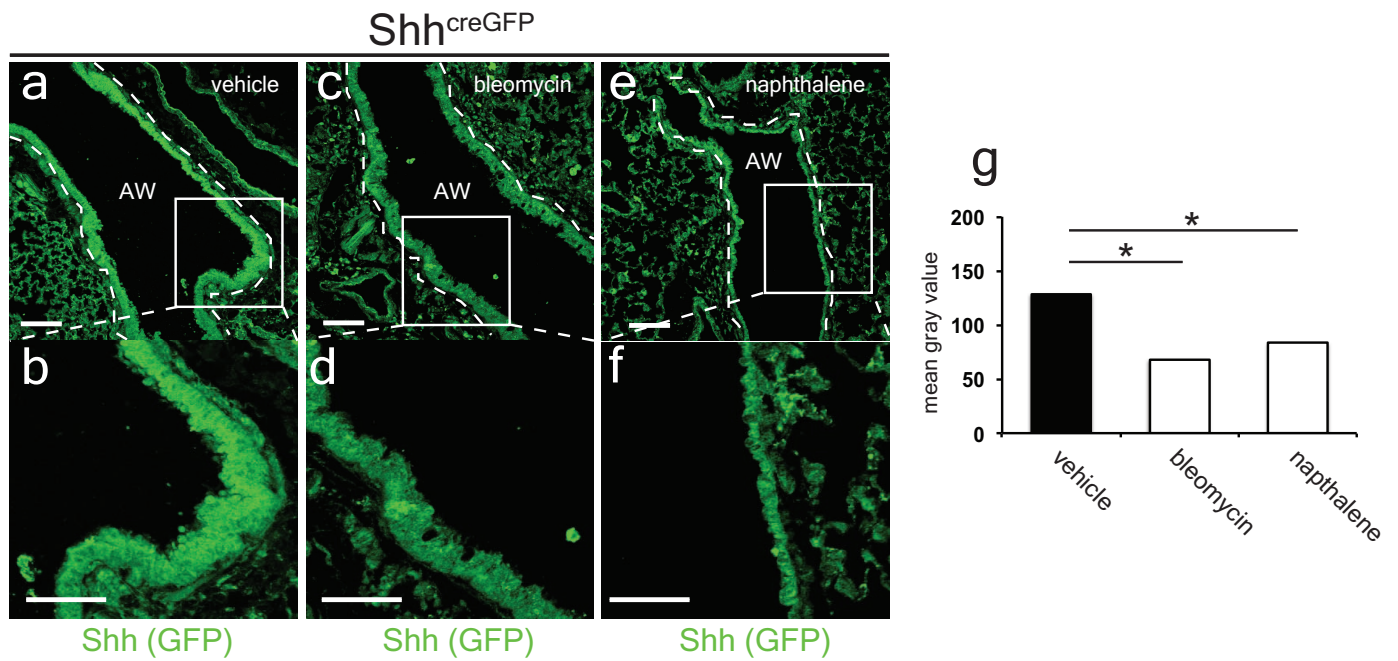
Extended Data Figure 5 | Shh expression is decreased with bleomycin and naphthalene injury to the airway. a–g, Repetitive bleomycin injury after one month or single-dose naphthalene injury after three days reduced GFP expression in the airways of the *Shh^creGFP* reporter compared to controls. Data represent $n = 2$ animals per group with 5 sections analysed per animal, AW, airway. Scale bars, 100 µm. *$P < 0.05$.
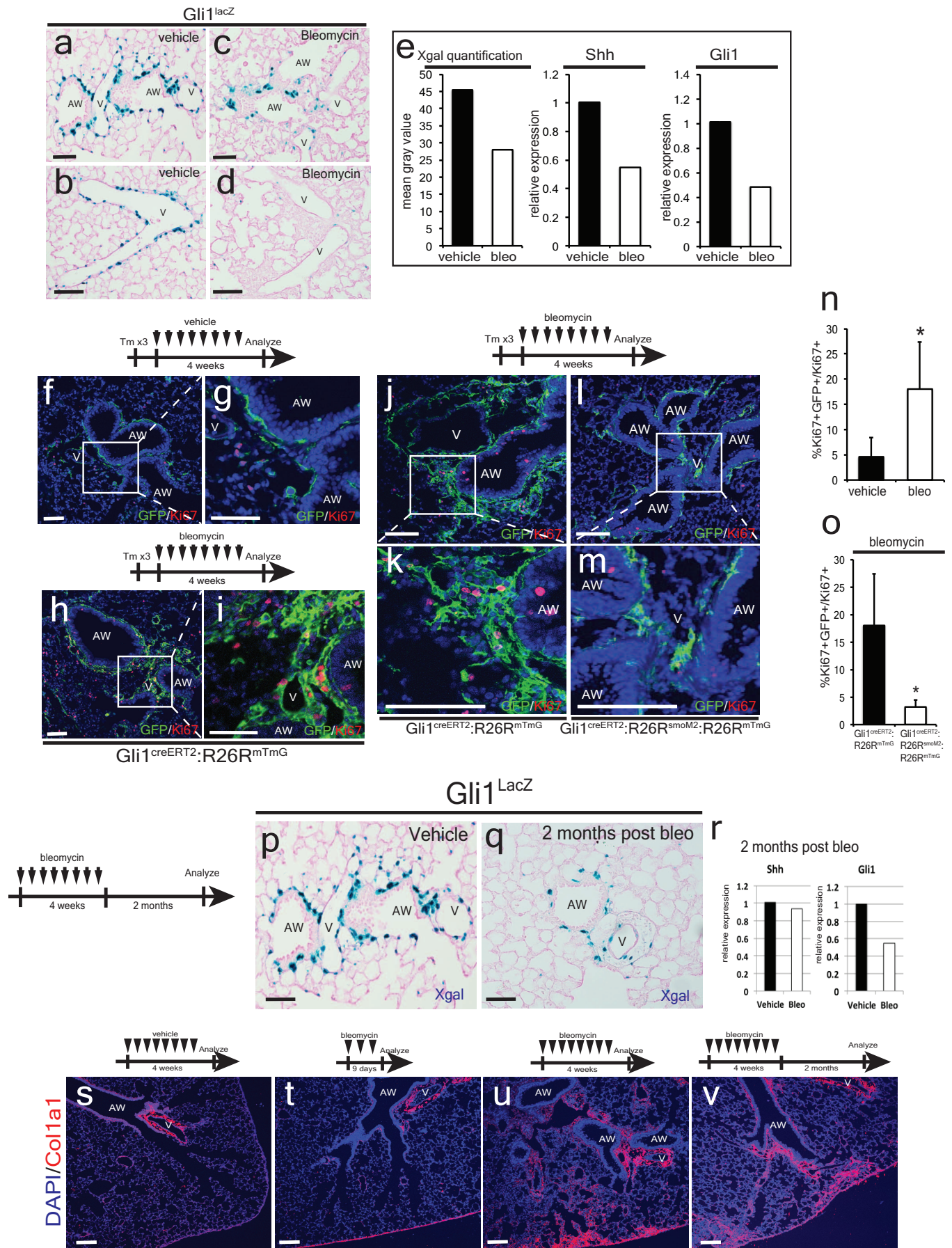
Extended Data Figure 6 | Hedgehog modulates mesenchymal response to bleomycin injury. **a–e**, Chronic repetitive injury to the lung epithelium with bleomycin over 4 weeks downregulates Gli1 expression in the mesenchyme adjacent to the airways in *Gli1^lacZ* lungs as noted by histochemical staining for β-galactosidase activity (**a–d**) and as noted using X-gal quantification and qPCR analysis of Shh and Gli1 expression after four weeks of repetitive injury (**e**, $n = 2$ animals per group). **f–i, n**, Lineage-traced, Gli1$^+$ Hh activated lung mesenchymal cells undergo proliferative expansion after repetitive bleomycin injury with an increase in Ki67$^+$ mesenchymal cells ($n = 4$ animals per group). **j–m, o**, Expression of Smo^{M2} within lineage traced Gli1$^+$ lung mesenchymal cells attenuates the proliferative expansion that normally follows repetitive bleomycin injury ($n = 3$ animals per group). **p–v**, Gli1 expression remains reduced 2 months after the end of bleomycin treatment (**p–r**, $n = 2$ animals per group), which might be due to the persistent scarring that is observed after repetitive bleomycin injury (**s–v**). AW, airway; V, blood vessel. Scale bars, 100 μm; *$P < 0.05$. Error bars, mean ± s.d.
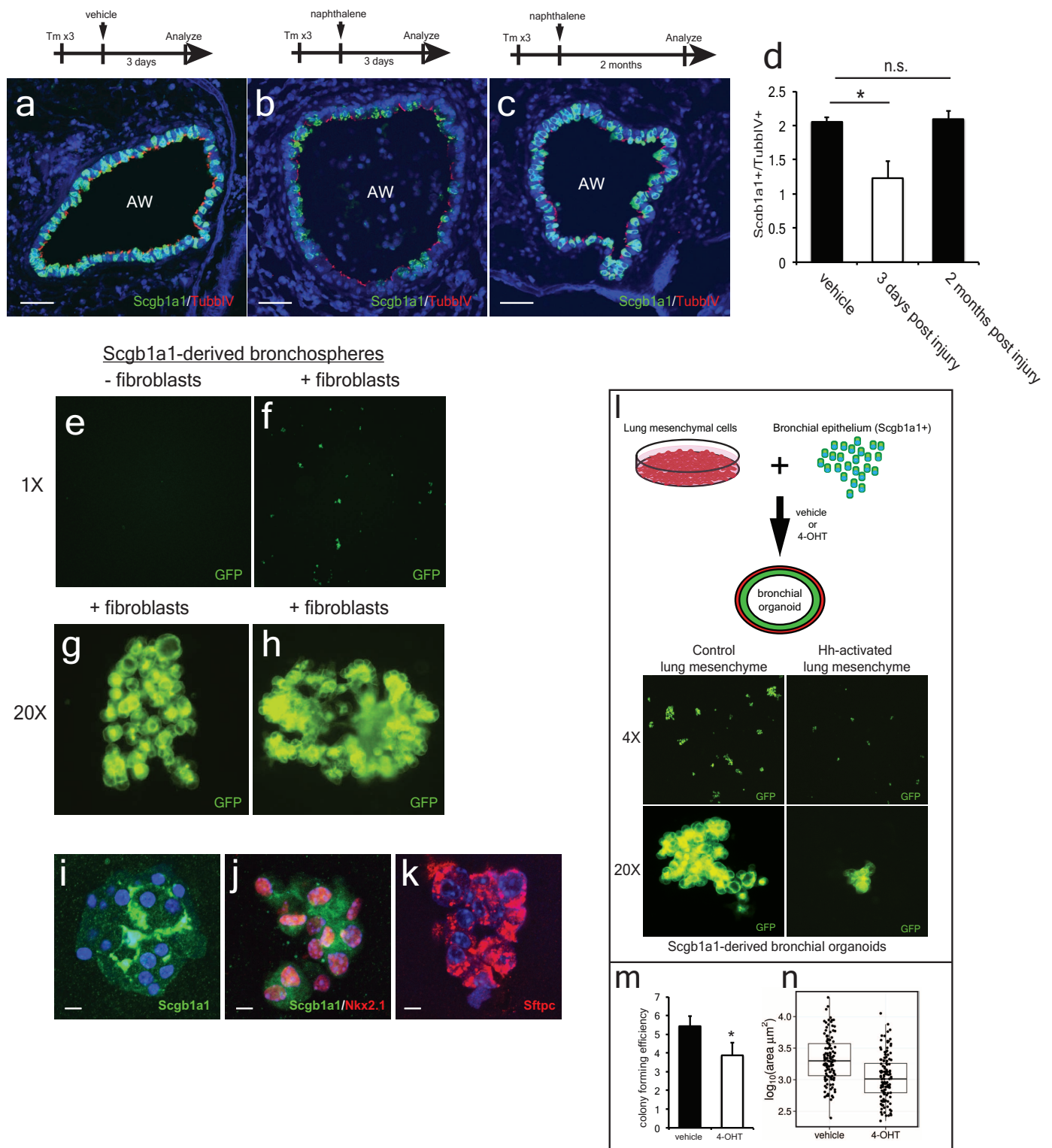
Extended Data Figure 7 | Airway epithelium is able to regenerate *in vitro* and *in vivo*. a, b, d, Scgb1a1⁺ secretory epithelium is initially depleted 3 days following naphthalene injury while TubbIV⁺ ciliated epithelium remains relatively intact (*n* = 3 animals per group). **c, d**, However, 2 months after the initial naphthalene injury, Scgb1a1⁺ secretory epithelium repopulates the airway and the ratio of secretory/ciliated cells is restored to levels before injury. **e, f**, GFP⁺ bronchial epithelial cells isolated from *Scgb1a1^{cre}:R26R^{mTmG}* animals were cultured in the presence or absence of isolated lung mesenchymal cells, and only those co-cultured with lung mesenchymal cells were able to form organoids. **g, h**, Examples of the 3-dimensional structures formed by the bronchial organoids in the presence of lung mesenchyme. **i–k**, Scgb1a1-derived organoids predominantly express markers of secretory airway differentiation, including Scgb1a1 and Nkx2.1 (**i, j**), while a minority expresses markers of alveolar epithelial lineage including Sftpc (**k**). **l–n**, Co-culture of lung mesenchyme and bronchial epithelium induces organoid formation (**l**), which is inhibited in number (**m**) and colony size (**n**) with activation of Hh in the mesenchyme. AW, airway. Scale bars, 100 μm. *$P < 0.05$. $n = 3$ animals per group for injury time points. Error bars, mean ± s.d. *In vitro* organoid studies represent technical quadruplicates, with > 140 randomly selected clones analysed for size per group.
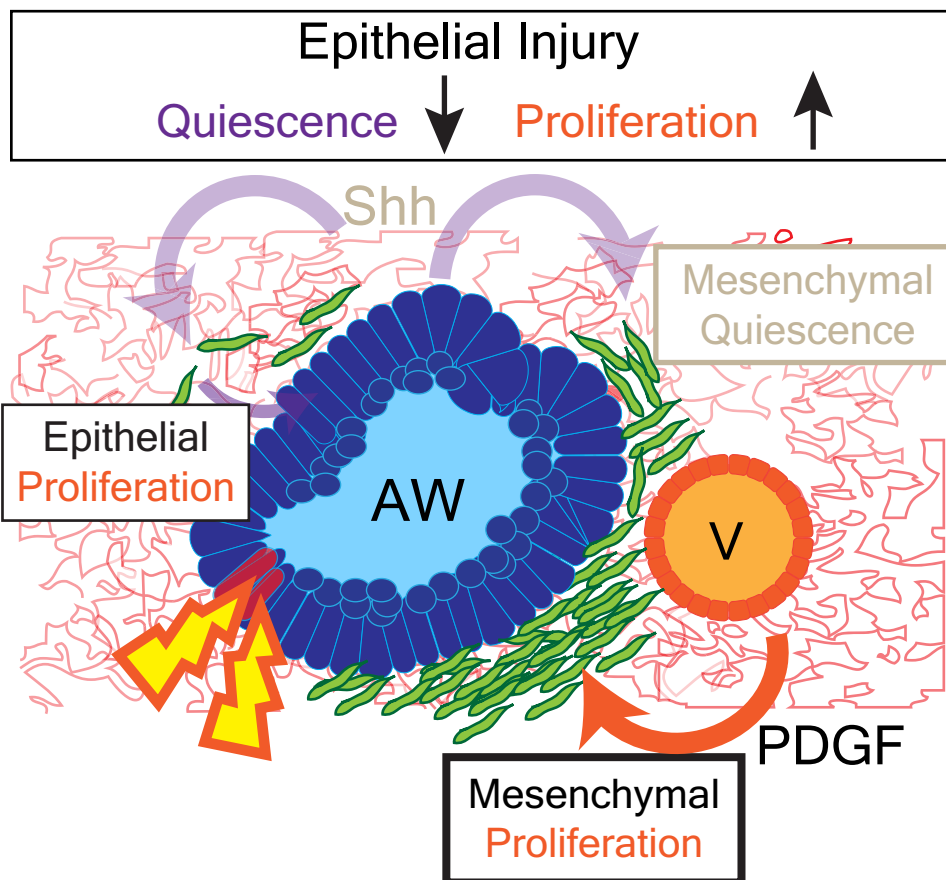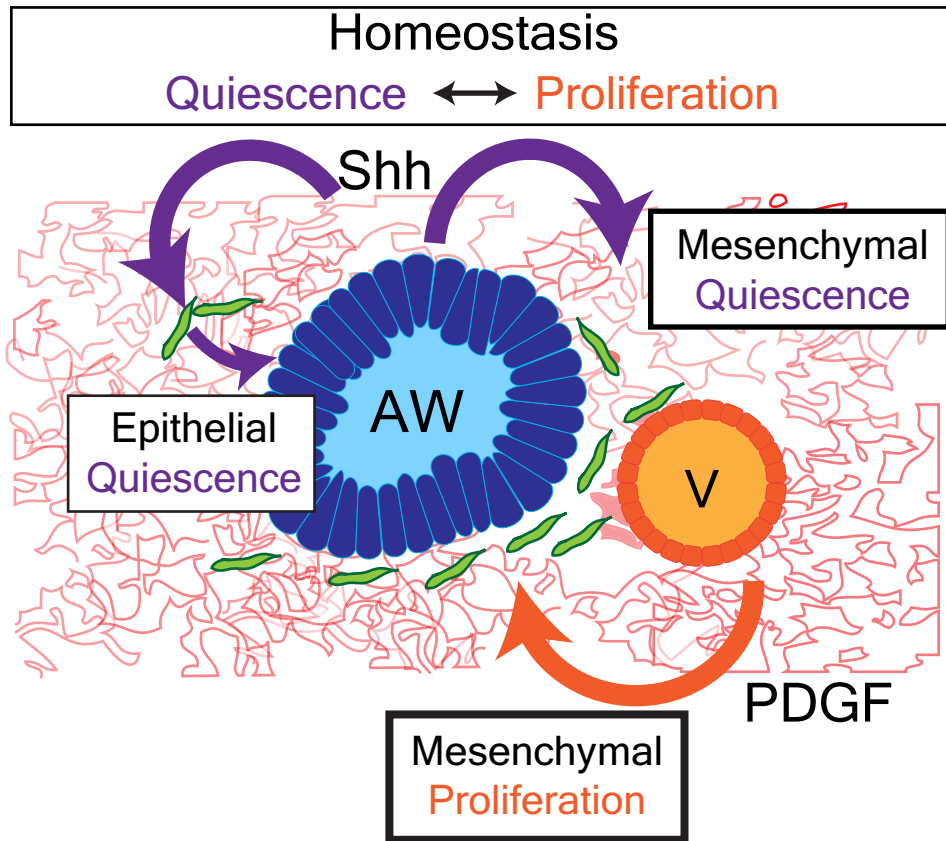
Extended Data Figure 8 | Hh signalling mediates both mesenchymal and epithelial quiescence during homeostasis and injury repair in the lung. The lung epithelium actively maintains mesenchymal quiescence through paracrine Hh signalling, which also regulates a feedback loop to maintain epithelial quiescence. Epithelial injury leads to downregulation of Hh signalling and loss of mesenchymal quiescence, which in turn stimulates epithelial regeneration to replete the airway epithelium until homeostasis is re-established.

LETTER

RAF inhibitors that evade paradoxical MAPK pathway activation

Chao Zhang[1], Wayne Spevak[1], Ying Zhang[1], Elizabeth A. Burton[1], Yan Ma[1], Gaston Habets[1], Jiazhong Zhang[1], Jack Lin[1], Todd Ewing[1], Bernice Matusow[1], Garson Tsang[1], Adhirai Marimuthu[1], Hanna Cho[1], Guoxian Wu[1], Weiru Wang[1], Daniel Fong[1], Hoa Nguyen[1], Songyuan Shi[1], Patrick Womack[1], Marika Nespi[1], Rafe Shellooe[1], Heidi Carias[1], Ben Powell[1], Emily Light[1], Laura Sanftner[1], Jason Walters[1], James Tsai[1], Brian L. West[1], Gary Visor[1], Hamid Rezaei[1], Paul S. Lin[1], Keith Nolop[1], Prabha N. Ibrahim[1], Peter Hirth[1] & Gideon Bollag[1]

Oncogenic activation of BRAF fuels cancer growth by constitutively promoting RAS-independent mitogen-activated protein kinase (MAPK) pathway signalling[1]. Accordingly, RAF inhibitors have brought substantially improved personalized treatment of metastatic melanoma[2–5]. However, these targeted agents have also revealed an unexpected consequence: stimulated growth of certain cancers[6–9]. Structurally diverse ATP-competitive RAF inhibitors can either inhibit or paradoxically activate the MAPK pathway, depending on whether activation is by BRAF mutation or by an upstream event, such as RAS mutation or receptor tyrosine kinase activation[10–12]. Here we have identified next-generation RAF inhibitors (dubbed 'paradox breakers') that suppress mutant BRAF cells without activating the MAPK pathway in cells bearing upstream activation. In cells that express the same HRAS mutation prevalent in squamous tumours from patients treated with RAF inhibitors, the first-generation RAF inhibitor vemurafenib stimulated _in vitro_ and _in vivo_ growth and induced expression of MAPK pathway response genes; by contrast the paradox breakers PLX7904 and PLX8394 had no effect. Paradox breakers also overcame several known mechanisms of resistance to first-generation RAF inhibitors. Dissociating MAPK pathway inhibition from paradoxical activation might yield both improved safety and more durable efficacy than first-generation RAF inhibitors, a concept currently undergoing human clinical evaluation with PLX8394.

Selective RAF inhibitors including vemurafenib[13] and dabrafenib[14] have demonstrated both objective tumour response and, in the case of vemurafenib, overall survival benefit in mutant BRAF[V600]-driven melanoma. The clinical effectiveness of RAF inhibitors depends on near-complete abolition of the MAPK pathway output in tumours harbouring BRAF mutations[13]. However, these compounds paradoxically activate

the MAPK pathway in cells bearing oncogenic RAS or elevated upstream receptor signalling[10–12]. This paradox can promote cellular proliferation and manifest clinically with progression of cutaneous squamous cell carcinomas (cuSCC) and keratoacanthomas, sometimes within weeks of therapy initiation[6,15]. These paradox-induced skin tumours have an uncharacteristically high incidence of RAS mutations[6,16], raising the concern that the same mechanism might accelerate progression of other RAS-driven cancers. Recent case reports of increased incidence of primary melanomas[7] and progression of RAS-mutant leukaemia and colon carcinoma during RAF inhibitor treatment[8,9] add weight to the concern. Although combination with MEK inhibition represents one strategy to combat paradoxical activation, and such combinations did show improved clinical responses[17,18], the combination of these two costly agents yields increased adverse events, and resistance still develops. Our strategy to develop next-generation RAF inhibitors is thus to design potent BRAF[V600] mutant inhibitors that avoid paradoxical activation of MAPK signalling.

Vemurafenib analogues with variable terminal sulfonamide and sulfamide substitutions were screened against a panel of cell lines for compound-induced change in phospho-ERK1/2 (T202/Y204, pERK). For each compound, the dissociation of pERK inhibition from activation (dubbed 'ERK pathway inhibition index' or EPII) was expressed as the ratio between the compound's mean pERK activation half-maximum effective concentration (EC$_{50}$) in three RAS mutant cell lines (murine cuSCC cell line B9, human melanoma cell line IPC-298, and human colorectal carcinoma cell line HCT116, Table 1), and the compound's mean pERK inhibition half-maximum inhibitory concentration (IC$_{50}$) in two BRAF[V600E] melanoma cell lines (A375 and COLO829, Table 1). The EPIIs for vemurafenib and dabrafenib were

Table 1 | Comparison of the _in vitro_ profile* of first-generation BRAF inhibitors with a paradox breaker

Compound	Biochemical IC$_{50}$ (µM)			pERK inhibition IC$_{50}$ (µM)		pERK activation EC$_{50}$ (µM)†			EPII‡
	BRAF[V600E]	BRAF	CRAF	A375	COLO829	B9	IPC-298	HCT116	
Vemurafenib	0.031 (±0.004)	0.1 (±0.02)	0.048 (±0.004)	0.032 (±0.007)	0.041 (±0.008)	0.36 (±0.08)	0.54 (±0.12)	0.34 (±0.07)	11
PLX4720	0.013 (±0.005)	0.16 (±0.03)	0.007 (±0.003)	0.044 (±0.006)	0.039 (±0.023)	0.24 (±0.03)	0.4 (±0.05)	0.29 (±0.17)	7
PLX7683	0.029 (±0.021)	1.1 (±0.6)	0.44 (±0.23)	0.98 (±0.75)	1.7 (±0.82)	>200	>200	>200	>100
PLX7904	0.0042 (±0.0006)	0.14 (±0.02)	0.091 (±0.014)	0.016 (±0.005)	0.018 (±0.005)	>200	>200	>200	>10,000
PLX8394	0.0038 (±0.0016)	0.014 (±0.004)	0.023 (±0.04)	0.0035 (±0.0012)	0.0021 (±0.0012)	>200	>200	>200	>50,000
PLX5568	0.58 (±0.07)	0.19 (±0.02)	0.021 (±0.002)	>10	>10	5.1 (±2.5)	3.2 (±1.9)	7 (±3.2)	<0.5
Sorafenib	0.35 (±0.04)	0.072 (±0.008)	0.011 (±0.002)	4.4 (±1.3)	2 (±1.2)	0.025 (±0.005)	0.019 (±0.01)	0.086 (±0.04)	0.01
Dabrafenib	0.0054 (±0.0015)	0.0027 (±0.001)	0.0015 (±0.001)	0.001 (±0.001)	0.005 (±0.003)	0.01§ (±0.005)	0.01§ (±0.005)	0.003§ (±0.002)	~4
PLX7922	0.012 (±0.008)	1.1 (±0.4)	0.053 (±0.006)	0.01 (±0.003)	0.014 (±0.008)	>10	3.3 (±2.4)	>10	>500

*Mutational status of the cell lines: A375, BRAF[V600E], homozygous; COLO829, BRAF[V600E], heterozygous; B9, HRAS[Q61L]; IPC-298, NRAS[Q61L]; HCT116, KRAS[G13D]. Each value is an average of more than four experiments. Values in parenthesis, s.e.m.
†EC$_{50}$, the concentration increasing pERK to 50% compared with the positive control, 10 µM PLX4720.
‡ERK pathway inhibition index (EPII), the ratio between mean pERK activation EC$_{50}$ and mean pERK inhibition IC$_{50}$.
§Using the rising portion of the concentration–response curve (Fig. 3c).

[1]Plexxikon Inc., 91 Bolivar Drive, Berkeley, California 94710, USA.

11 and 4, respectively. Among the compounds that exhibited more than 100-fold EPII was a molecule (PLX7683, Fig. 1a and Table 1) that contained an N-ethylmethyl-sulfamide moiety in lieu of the propyl-sulfonamide tail of vemurafenib. Optimization of this series of compounds by substitution on the 5-position of the 7-azaindole scaffold generated PLX7904 (Fig. 1a), which potently inhibited pERK in BRAF^{V600E} cells but showed essentially no pERK activation in RAS mutant cell lines at the concentrations tested (Table 1, Fig. 1b and Extended Data Fig. 1). PLX7904 was also evaluated in the human SCC cell line A431 and the human breast adenocarcinoma cell line SKBR3 as these cells achieve MAPK pathway activation by upstream signals feeding into RAS (through overexpression of epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2), respectively). Unlike vemurafenib, PLX7904 did not increase pERK levels in these cells (Fig. 1c). In biochemical assays using recombinant kinases, PLX7904 showed preferential inhibition of the mutated BRAF^{V600E} over wild-type

BRAF and CRAF and a level of kinome selectivity comparable to that of vemurafenib[13] (Supplementary Table 1).

PLX7904 inhibited the *in vitro* growth of two aforementioned melanoma cell lines (A375 and COLO829) and an additional human colorectal cancer cell line COLO205 that expressed BRAF^{V600E} with IC$_{50}$ values of 0.17 μM, 0.53 μM, and 0.16 μM, respectively, on a par with vemurafenib IC$_{50}$ values in the same assays (0.33 μM, 0.69 μM, and 0.25 μM, respectively). Consistent with this *in vitro* result, PLX7904 and vemurafenib produced similar anti-tumour effects in a subcutaneous COLO205 xenograft model (Fig. 1e) with matching doses (25 mg per kg twice daily) and plasma exposures (steady-state area under the curve ≈ 200,000 ng ml^{-1} h).

Recent analyses[6,16] of the cuSCC and keratoacanthoma lesions excised from vemurafenib recipients revealed that up to 60% of the specimens harboured RAS mutations, mostly HRAS^{Q61L}, supporting an important role of RAS mutation in BRAF inhibitor-induced cuSCC. The B9 cuSCC mouse cell line expresses the same activated HRAS^{Q61L} allele[19]. In soft agar, both vemurafenib and its analogue PLX4720 stimulated B9 colony formation at concentrations similar to the growth inhibitory IC$_{50}$ values in A375, COLO829, and COLO205 cells, whereas PLX7904 did not (Fig. 1d). When tested *in vivo*, subcutaneous B9-tumour growth was accelerated by vemurafenib but not by the equally potent BRAF^{V600E} inhibitor PLX7904 when administered at the same dose (Fig. 1f).

We compared the gene expression changes in B9 cells treated overnight with vemurafenib and PLX7904. Vemurafenib altered transcription of 191 mouse genes by at least 1.9-fold, while PLX7904 had minimal effects (Extended Data Fig. 2 and Supplementary Table 2). Of the genes significantly induced by vemurafenib, three encode EGFR ligands: amphiregulin, heparin-binding EGF-like growth factor, and transforming growth factor-α (TGF-α) (Extended Data Fig. 2c). The upregulation of these autocrine growth factors was confirmed at protein level and their role in potentiating the transforming potential of activated HRAS was demonstrated (Extended Data Figs 3 and 4). Induction of these ligands by vemurafenib has been demonstrated independently in vemurafenib-resistant lung cancer cell lines[20]. All three ligands can promote cuSCC[21]. These data implicate EGFR signalling as a potential molecular link between paradoxical MAPK activation by RAF inhibitors and secondary malignancies. In contrast to vemurafenib and consistent with the paradox breaker profile, expression of the EGFR ligands was largely unaffected by PLX7904 (Extended Data Figs 2–4).

PLX7904 and a further optimized analogue PLX8394 (Fig. 1a and Table 1) are only subtly different from vemurafenib based on chemical structure. To understand how such small molecular alterations cause a drastic change in the biological profile, we obtained the crystal structure of PLX7904 in complex with BRAF^{V600E} (Extended Data Table 1). The overall binding of PLX7904 (Fig. 2a) is similar to that of vemurafenib (Extended Data Fig. 5a) with the terminal N-ethylmethyl group of PLX7904 occupying the same small interior pocket as the propyl group of vemurafenib. However, the methyl group of the N-ethylmethyl moiety forms closer contact with Leu505 in the pocket (Fig. 2b). Leu505 is one of the four residues that compose the so-called regulatory spine of kinases[22] (Extended Data Fig. 6). Situated close to the carboxy (C)-terminal end of the αC helix, Leu505 is the only residue from that helix that makes direct contact with the inhibitor. RAS promotes RAF dimerization, and paradoxical MAPK pathway activation results from binding of the inhibitor to one protomer of a RAF dimer which allosterically transactivates the other protomer[10–12]. The αC helix plays a critical role in RAF dimer formation[10,23] and mutations that disrupt the αC helix dimer contacts counteract RAF activation by inhibitors. In an enzyme-linked immunosorbent assay (ELISA) of dimerization using cell lysates, vemurafenib and other known BRAF inhibitors promote BRAF–CRAF heterodimer formation in RAS mutant cells, whereas the dimer formation is indifferent to the presence of PLX7904 (Fig. 2c). Although the crystal structure did
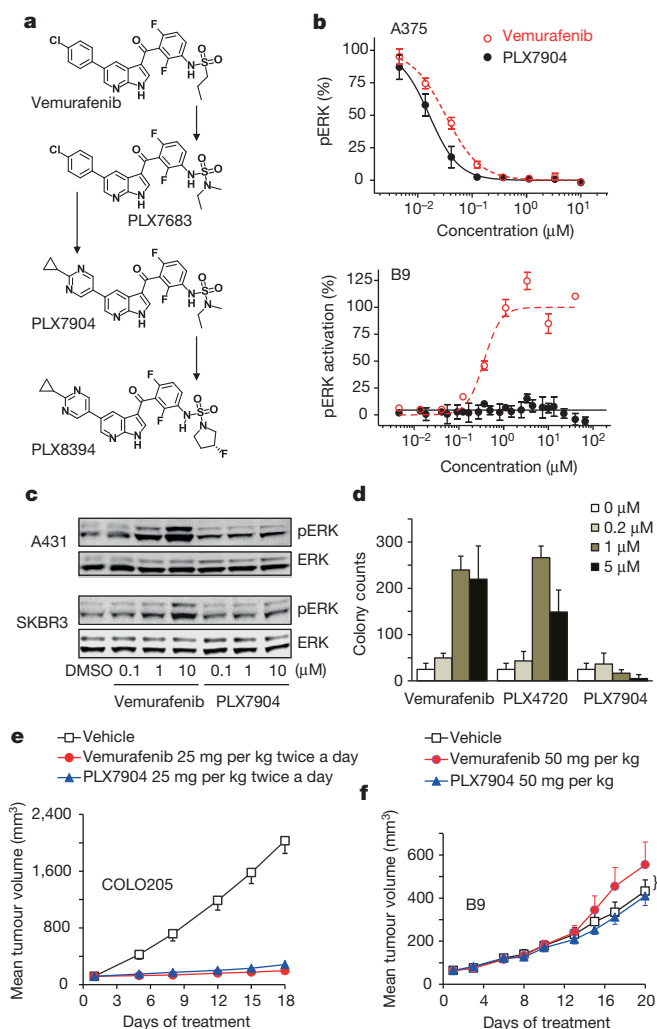
Figure 1 | Paradox breakers dissociate MAPK pathway inhibition from activation. a, Vemurafenib and paradox breakers PLX7683, PLX7904, and PLX8394. **b**, pERK IC$_{50}$ curves (mean ± s.d.) in A375 (BRAF^{V600E}) cells and pERK EC$_{50}$ curves (mean ± s.d., normalized to maximal pERK level induced by PLX4720) in B9 (HRAS^{Q61L}) cells (n = 5 experiments). **c**, pERK in A431 and SKBR3 cells after treatment for 1 h with vemurafenib or PLX7904 (full scans of western blot in Supplementary Figure 1). Repeated three times. **d**, Anchorage-independent growth of B9 cells with vemurafenib and PLX4720 (for 3 weeks) but not PLX7904 (mean ± s.d., two experiments, three replicates each). **e**, PLX7904 and vemurafenib inhibited the COLO205 xenograft growth (mean ± s.e.m., eight mice per group). **f**, B9 subcutaneous xenografts were stimulated by vemurafenib (*P < 0.05 by two-sided t-test) but not by PLX7904 (mean ± s.e.m., ten mice per group).

Figure 2 | Molecular mechanisms of paradox breakers. a, Interactions between PLX7904 (green) and BRAF^{V600E} (grey). Phe595 (spheres) shows the DFG-in (or type 1) conformation. Red dashed lines represent hydrogen bonds. **b**, The propyl-sulfonamide tail of vemurafenib (cyan) and the N-ethylmethyl-sulfamide tail of PLX7904 (green) viewed from the dimer interface. Four residues (Ile527, Leu505, Phe595, and Leu567) form the R-spine (see Extended Data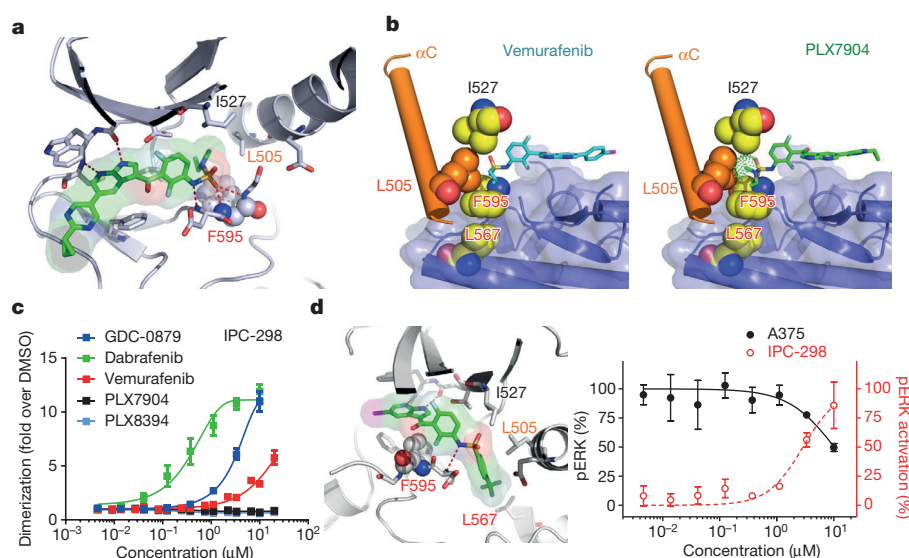 Fig. 6 for definition). A dotted surface around the N-methyl group in PLX7904 illustrates its close contact with Leu505 from the αC helix (orange). **c**, BRAF–CRAF heterodimer formation in IPC-298 cells with increasing concentrations of RAF inhibitors (treatment for 1 h; mean ± s.d., $n = 5$ experiments). **d**, Type 2 (that is, DFG-out) binder PLX5568 is a CRAF-selective inhibitor with inverse EPII (mean ± s.d., $n = 5$ experiments).

not reveal further displacement of the αC helix by PLX7904 (Extended Data Fig. 5c), compared with the vemurafenib-bound structure the three terminal atoms (Cγ, Cδ1, Cδ2) of Leu505 shifted by 0.6–1 Å to accommodate PLX7904 (Extended Data Fig. 5d). Leu505 has a higher side-chain crystallographic temperature factor (B-factor = 75) than the same residue in the vemurafenib structure (B-factor = 33). In solution where the protein is free of the artificial constraints of crystal, the strong interaction between PLX7904 and Leu505 could lead to outward movement of the αC helix, causing disruption to the RAF dimer interface. The vemurafenib-resistant L505H BRAF mutant remained sensitive to PLX7904 (ref. 24), supporting the key role played by residue 505 in sensing the structural difference between paradox breakers and first-generation RAF inhibitors.

The crystal structure of PLX4720 in complex with wild-type BRAF showed that the compound adopts a type 2 kinase inhibitor binding pose when accessing the inactive conformation of the kinase, the preferred state of wild-type RAF proteins[25]. A PLX4720 analogue, PLX5568, made to enforce the type 2 binding orientation (Fig. 2d), has intrinsic selectivity towards CRAF (Table 1). Like other type 2 RAF inhibitors such as sorafenib, PLX5568 showed marginal inhibitory activity against BRAF^{V600E} cells but still paradoxically activated

pERK in mutant RAS cells, thus exhibiting inverse EPII (Table 1 and Fig. 2d). These data highlight the existence of a strong correlation between conformation-specific inhibition and biological outcome.

A newly discovered[26] and potentially common[27] mechanism whereby BRAF^{V600E} melanomas develop resistance to BRAF inhibition is to express aberrantly spliced forms of BRAF^{V600E} that can dimerize in the absence of activated RAS. Wild-type BRAF can be activated in a similar manner when a chromosomal translocation event results in a truncated C-terminal fragment of BRAF embedded in a fusion gene with oncogenic activity (Supplementary Table 3 and references therein). The fusion kinase, like the spliced forms of BRAF^{V600E}, dimerizes and has constitutive kinase activity and intrinsic resistance to first-generation RAF inhibitors. To test whether the paradox breaker strategy can be exploited to combat dimerization-mediated resistance, we compared the activity with PLX7904 in the SK-MEL-239 parental cell line and a representative vemurafenib-resistant clone (C3) expressing a truncated BRAF^{V600E}. PLX7904 demonstrated minimal shift in pMEK IC$_{50}$ and modest increase in growth inhibition IC$_{50}$ in C3 cells (Extended Data Fig. 7). Furthermore, both PLX7904 (that is, PB04) and PLX8394 (that is, PB03) overcame RAF inhibitor resistance in BRAF fusions characterizing paediatric astrocytomas[28] and maintained
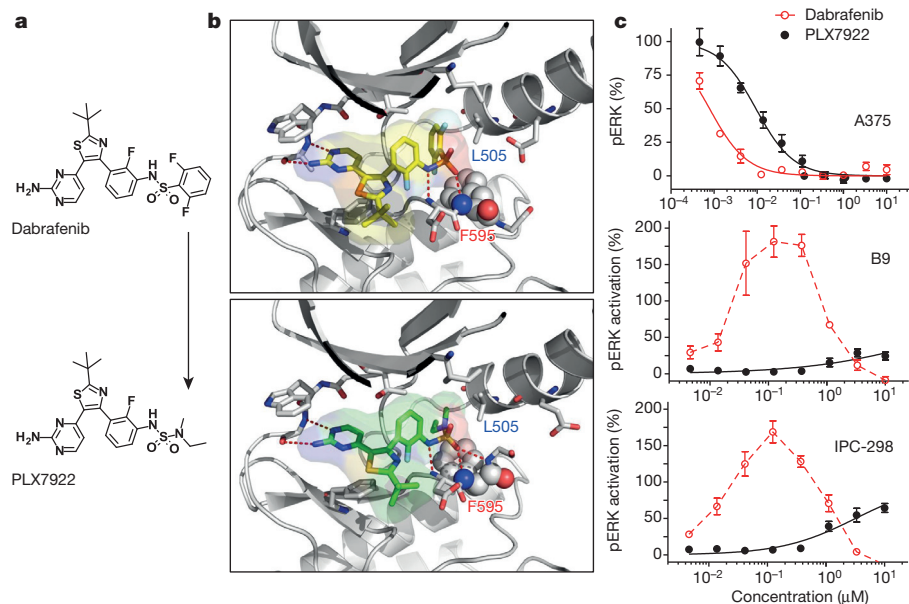
Figure 3 | The structural determinant of paradox breakers can be transferred to another chemical series to drastically alter its biological profile. a, Substituting the 2,6-difluoro-phenylsulfonamide of dabrafenib with an N-ethylmethyl-sulfamide tail yielded PLX7922. **b**, Crystal structures of dabrafenib and PLX7922 in complex with BRAF^{V600E}. **c**, Dabrafenib (red), a highly potent inhibitor of pERK in BRAF^{V600E} cell lines, exhibited an unusual bell-shaped pERK activation curve in mutant NRAS cell lines (B9 and IPC-298). PLX7922 (black) shows markedly reduced pERK activation in mutant RAS cells with only moderately decreased pERK IC$_{50}$ in BRAF^{V600E} cells (A375). Mean ± s.d. in **c**, $n = 5$ experiments.

activity against cells that are vemurafenib-resistant through secondary mutation in NRAS[29].

The discovery of paradox breakers confirms that the two opposing modes of action of RAF inhibitors, either blocking or activating the MAPK pathway, can be uncoupled. Since the tail moiety is primarily responsible for this uncoupling, we engineered the PLX7904 tail onto the dabrafenib scaffold (Fig. 3a, b). The resulting compound, PLX7922, showed significantly increased EPII (Fig. 3c and Table 1). Thus, the structural and chemical principles of paradox breaking can be applied to improve the safety and biological profile of other RAF inhibitors.

An alternative strategy to overcome paradoxical activation is to completely block all RAF isoforms (pan-RAF inhibition), thus severing the link between RAS and MEK/ERK. AZ-628, the first compound in this class, did show reduced (rather than induced) pERK/pMEK in RAS mutant cells, but possesses unfavourable pharmaceutical properties[10]. Recently, new pan-RAF inhibitors with ancillary activity on upstream SRC family kinases have been reported[30]. The concern for pan-RAF inhibitors is that blocking MAPK signalling in normal tissue will cause toxicity. Thus, paradox breakers should afford a higher therapeutic index than the first-generation RAF inhibitors and pan-RAF inhibitors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Davies, H. et al. Mutations of the BRAF gene in human cancer. Nature 417, 949–954 (2002).
2. Flaherty, K. T. et al. Inhibition of mutated, activated BRAF in metastatic melanoma. N. Engl. J. Med. 363, 809–819 (2010).
3. Chapman, P. B. et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N. Engl. J. Med. 364, 2507–2516 (2011).
4. Sosman, J. A. et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. N. Engl. J. Med. 366, 707–714 (2012).
5. Hauschild, A. et al. Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. Lancet 380, 358–365 (2012).
6. Su, F. et al. RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. N. Engl. J. Med. 366, 207–215 (2012).
7. Zimmer, L. et al. Atypical melanocytic proliferations and new primary melanomas in patients with advanced melanoma undergoing selective BRAF inhibition. J. Clin. Oncol. 30, 2375–2383 (2012).
8. Callahan, M. K. et al. Progression of RAS-mutant leukemia during RAF inhibitor treatment. N. Engl. J. Med. 367, 2316–2321 (2012).
9. Andrews, M. C. et al. BRAF inhibitor-driven tumor proliferation in a KRAS-mutated colon carcinoma is not overcome by MEK1/2 inhibition. J. Clin. Oncol. 31, e448–e451 (2013).
10. Hatzivassiliou, G. et al. RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. Nature 464, 431–435 (2010).
11. Heidorn, S. J. et al. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. Cell 140, 209–221 (2010).
12. Poulikakos, P. I., Zhang, C., Bollag, G., Shokat, K. M. & Rosen, N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature 464, 427–430 (2010).
13. Bollag, G. et al. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature 467, 596–599 (2010).
14. King, A. J. et al. Dabrafenib; preclinical characterization, increased efficacy when combined with trametinib, while BRAF/MEK tool combination reduced skin lesions. PLoS One 8, e67583 (2013).
15. Anforth, R. M. et al. Cutaneous manifestations of dabrafenib (GSK2118436): a selective inhibitor of mutant BRAF in patients with metastatic melanoma. Br. J. Dermatol. 167, 1153–1160 (2012).
16. Oberholzer, P. A. et al. RAS mutations are associated with the development of cutaneous squamous cell tumors in patients treated with RAF inhibitors. J. Clin. Oncol. 30, 316–321 (2012).
17. Larkin, J. et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. N. Engl. J. Med. 371, 1867–1876 (2014).
18. Robert, C. et al. Improved overall survival in melanoma with combined dabrafenib and trametinib. N. Engl. J. Med. 372, 30–39 (2015).
19. Balmain, A., Ramsden, M., Bowden, G. T. & Smith, J. Activation of the mouse cellular Harvey-ras gene in chemically induced benign skin papillomas. Nature 307, 658–660 (1984).
20. Lin, L. et al. Mapping the molecular determinants of BRAF oncogene dependence in human lung cancer. Proc. Natl Acad. Sci. USA 111, E748–E757 (2014).
21. Oshima, G. et al. Autocrine epidermal growth factor receptor ligand production and cetuximab response in head and neck squamous cell carcinoma cell lines. J. Cancer Res. Clin. Oncol. 138, 491–499 (2012).
22. Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem. Sci. 36, 65–77 (2011).
23. Freeman, A. K., Ritt, D. A. & Morrison, D. K. Effects of Raf dimerization and its inhibition on normal and disease-associated Raf signaling. Mol. Cell 49, 751–758 (2013).
24. Choi, J. et al. Identification of PLX4032-resistance mechanisms and implications for novel RAF inhibitors. Pigment Cell Melanoma Res. 27, 253–262 (2014).
25. Tsai, J. et al. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. Proc. Natl Acad. Sci. USA 105, 3041–3046 (2008).
26. Poulikakos, P. I. et al. RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). Nature 480, 387–390 (2011).
27. Rizos, H. et al. BRAF inhibitor resistance mechanisms in metastatic melanoma: spectrum and clinical impact. Clin. Cancer Res. 20, 1965–1977 (2014).
28. Sievert, A. J. et al. Paradoxical activation and RAF inhibitor resistance of BRAF protein kinase fusions characterizing pediatric astrocytomas. Proc. Natl Acad. Sci. USA 110, 5957–5962 (2013).
29. Le, K., Blomain, E. S., Rodeck, U. & Aplin, A. E. Selective RAF inhibitor impairs ERK1/2 phosphorylation and growth in mutant NRAS, vemurafenib-resistant melanoma cells. Pigment Cell Melanoma Res. 26, 509–517 (2013).
30. Girotti, M. R. et al. Paradox-breaking RAF inhibitors that also target SRC are effective in drug-resistant BRAF mutant melanoma. Cancer Cell 27, 85–96 (2015).

METHODS

With the exception of tumour xenograft studies, no statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Generation of BRAF small molecule inhibitors. All solvents and reagents were used as obtained from commercial sources. Starting materials were purchased from commercial sources or prepared according to methods reported in the literature. Reactions involving air- or moisture-sensitive reagents were performed under a nitrogen atmosphere. NMR spectra were recorded in deuterated solvent with an Agilent 400 MHz MR DD2 spectrometer system equipped with an Oxford AS400 magnet. Chemical shifts are expressed as δ units and referenced to the residual ^{1}H or ^{13}C solvent signal. All coupling constants (J) are reported in hertz (s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; br, broad peak; dd, doublet of doublets; ddd, doublet of doublet of doublets; dm, doublet of multiplets). Mass spectra were measured with a Shimadzu LCMS-2020 spectrometer coupled to a Shimadzu 20A high-performance liquid chromatography (HPLC) system operating in reverse mode. Analytical purity was greater than 95% for final compounds and was determined using the following HPLC method. Buffer A: 5% acetonitrile, 95% water, 0.01% formic acid; buffer B: 95% CH_3CN, 5% water, 0.01% formic acid; SiliaChrom XDB C18, 5 μm, 2.1 × 50 mm, 5–95% B in 6 min, 1.0 ml min^{-1}, 220 nm and 254 nm, electrospray-ionization-positive (ESI-positive), 300–800 atomic mass units.

(2,6-Difluoro-3-nitrophenyl)(5-iodo-1H-pyrrolo[2,3-b]pyridin-3-yl)methanone. 5-Iodo-1H-pyrrolo[2,3-b]pyridine (160 g, 0.656 mol) and aluminium chloride (525 g, 3.94 mol) in nitromethane (1,640 ml) were allowed to stir at room temperature (20–25 °C) for 1 h. Then 2,6-difluoro-3-nitrobenzoyl chloride (218 g, 0.984 mmol) in nitromethane (1,640 ml) was added and the mixture was heated at 50 °C for 4 days. After cooling to 0 °C, the reaction was quenched with the dropwise addition of methanol (1.5 l), resulting in a precipitate. The mixture was diluted with water (2 l) and filtered. The crude product was triturated with methyl tert-butyl ether and filtered to give the title compound as a tan solid which was used directly in the next step (281 g, theory) without further purification. ^{1}H NMR (400 MHz, dimethylsulfoxide (DMSO)-d$_6$) 13.18 (br s, 1 H), 8.82 (s, 1 H), 8.62 (s, 1 H), 8.46 (m, 1 H), 8.40 (s, 1 H), 7.55 (m, 1 H).

(3-Amino-2,6-difluorophenyl)(5-iodo-1H-pyrrolo[2,3-b]pyridin-3-yl)methanone. To (2,6-difluoro-3-nitrophenyl)(5-iodo-1H-pyrrolo[2,3-b]pyridin-3-yl)methanone (281 g, 656 mmol) in ethyl acetate (10.9 l) and tetrahydrofuran (10.9 l) was added tin(II) chloride dihydrate (517 g, 2.29 mol) portionwise while heating at 60 °C. The reaction mixture was held at this temperature overnight. After cooling to room temperature, the reaction mixture was quenched with 50% saturated aqueous sodium bicarbonate (1:1 water and saturated aqueous sodium bicarbonate) and filtered through Celite washing the cake with ethyl acetate. The layers were separated and the organic layer was washed with brine and then concentrated under reduced pressure to give the crude product, which was triturated with methyl tert-butyl ether and filtered to give the title compound as a tan solid (216 g, 541 mmol, 83% yield). ^{1}H NMR (400 MHz, DMSO-d$_6$) 12.96 (br s, 1 H), 8.72 (s, 1 H), 8.56 (d, J = 2.0 Hz, 1 H), 8.06 (s, 1 H), 6.92 (dd, J = 8.6 Hz, 1 H), 6.88 (m, 1 H), 5.20 (s, 2 H); liquid chromatography–mass spectrometry (LC/MS) (ESI-positive) m/z: 399.9 (M + H$^+$).

(3-Amino-2,6-difluorophenyl)(5-(2-cyclopropylpyrimidin-5-yl)-1H-pyrrolo-[2,3-b]pyridin-3-yl)methanone. A mixture of (3-amino-2,6-difluorophenyl)(5-iodo-1H-pyrrolo[2,3-b]pyridin-3-yl)methanone (93 g, 233 mmol), 2-cyclopropyl-5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyrimidine (229 g, 466 mmol, ~ 50% purity), potassium carbonate (97.0 g, 702 mmol), and [1,1′-bis-(diphenylphosphino)ferrocene]dichloropalladium(II) dichloromethane complex (19.0 g, 23.3 mmol) in dioxane (930 ml) and water (465 ml) was heated at 100 °C for several hours. Upon cooling, the reaction mixture was diluted with water and extracted with a mixture of tetrahydrofuran and ethyl acetate. The organic layer was separated and concentrated under reduced pressure to give the crude product, which was triturated with dichloromethane/methyl tert-butyl ether and filtered, washing with methyl tert-butyl ether to give the title compound as a tan solid (71.0 g, 78% yield). ^{1}H NMR (400 MHz, DMSO-d$_6$) 12.95 (br s, 1 H), 9.07 (s, 2 H), 8.71 (d, J = 2.3 Hz, 1 H), 8.66 (s, 1 H), 8.11 (s, 1 H), 6.92 (dd, J = 9.0 Hz, 9.0 Hz, 1 H), 6.89 (ddd, J = 5.9 Hz, 9.0 Hz, 9.0 Hz, 1 H), 5.20 (s, 2 H), 2.27 (m, 1 H), 1.03–1.22 (m, 4 H); LC/MS (ESI-positive) m/z: 392.2 (M + H$^+$).

5-(2-Cyclopropylpyrimidin-5-yl)-3-[3-[[ethyl(methyl)sulfamoyl]amino]-2,6-difluoro-benzoyl]-1H-pyrrolo[2,3-b]pyridine (PLX7904). To (3-amino-2,6-difluorophenyl)(5-(2-cyclopropylpyrimidin-5-yl)-1H-pyrrolo[2,3-b]pyridin-3-yl)methanone (53.8 g, 138 mmol) in pyridine (1375 ml) was added ethyl(methyl)sulfamoyl chloride (65.0 g, 412 mmol) and the reaction was heated at 65 °C overnight. The volatiles were removed under reduced pressure and the residue was partitioned between water and ethyl acetate/tetrahydrofuran. The organic layer was

concentrated under reduced pressure to give the crude product, which was dry loaded onto silica gel and purified by silica gel column chromatography (twice) eluting with 0–10% methanol/dichloromethane, then purified by silica gel column chromatography eluting with ethyl acetate. The fractions containing the desired product were pooled and concentrated under reduced pressure. The resulting solid was triturated with methyl tert-butyl ether and filtered to give the title compound as a white solid (21.1 g, 30% yield). ^{1}H NMR (400 MHz, DMSO-d$_6$) 13.07 (br s, 1 H), 9.71 (br s, 1 H), 9.03 (s, 2 H), 8.76 (s, 1 H), 8.68 (s, 1 H), 8.19 (s, 1 H), 7.59 (ddd, J = 5.9 Hz, 9.0 Hz, 9.0 Hz, 1 H), 7.27 (dd, J = 9.0 Hz, 9.0 Hz, 1 H), 3.12 (q, J = 7.0 Hz, 2 H), 2.74 (s, 3 H), 2.29 (m, 1 H), 1.03–1.22 (m, 4 H), 0.95 (t, J = 7.0 Hz, 3 H); ^{13}C NMR (100 M Hz, DMSO-d6) 181.1, 170.4, 156.1 (dd, J$_{CF}$ = 246 Hz, J$_{CF}$ = 6.9 Hz), 155.5, 152.4 (dd, J$_{CF}$ = 250 Hz, J$_{CF}$ = 8.4 Hz), 149.7, 144.3, 139.2, 128.9, 128.6 (d, J$_{CF}$ = 9.9 Hz), 127.7, 126.2, 123.0 (dd, J$_{CF}$ = 13.3 Hz, J$_{CF}$ = 3.4 Hz), 118.5 (dd, J$_{CF}$ = 24.6 Hz, J$_{CF}$ = 22.5 Hz), 117.9, 116.2, 112.7 (dd, J$_{CF}$ = 22.5 Hz, J$_{CF}$ = 3.4 Hz), 45.3, 34.4, 18.2, 13.3, 10.9; LC/MS (ESI-positive) m/z: 513.3 (M + H$^+$).

(3R)-N-[3-[5-(2-cyclopropylpyrimidin-5-yl)-1H-pyrrolo[2,3-b]pyridine-3-carbonyl]-2,4-difluoro-phenyl]-3-fluoro-pyrrolidine-1-sulfonamide (PLX8394). This material was prepared in a manner analogous to PLX7904 using (3R)-3-fluoropyrrolidine-1-sulfonyl chloride in place of ethyl(methyl)sulfamoyl chloride. The material was purified by reverse-phase HPLC to provide, after lyophilization, the title compound as a white solid. ^{1}H NMR (400 MHz, DMSO-d$_6$) 13.05 (br s, 1 H), 9.84 (br s, 1 H), 9.01 (s, 2 H), 8.73 (s, 1 H), 8.67 (s, 1 H), 8.15 (s, 1 H), 7.62 (ddd, J = 5.9 Hz, 9.0 Hz, 9.0 Hz, 1 H), 7.26 (dd, J = 9.0 Hz, 9.0 Hz, 1 H), 5.29 (dm, J = 51.6 Hz (H-F), 1 H), 3.43 (dm, 2 H), 3.33 (m, 2 H), 2.27 (m, 1 H), 2.04 (m, 2 H), 1.01–1.11 (m, 4 H); ^{13}C NMR (100 MHz, DMSO-d6) 181.1, 170.4, 156.2 (dd, J$_{CF}$ = 247 Hz, J$_{CF}$ = 6.9 Hz), 155.5, 152.6 (dd, J$_{CF}$ = 249 Hz, J$_{CF}$ = 8.4 Hz), 149.7, 144.3, 139.2, 128.9, 128.7 (d, J$_{CF}$ = 9.2 Hz), 127.7, 126.2, 122.9 (dd, J$_{CF}$ = 13.7 Hz, J$_{CF}$ = 3.8 Hz), 118.5 (dd, J$_{CF}$ = 24.4 Hz, J$_{CF}$ = 22.2 Hz), 117.9, 116.2, 112.7 (dd, J$_{CF}$ = 22.9 Hz, J$_{CF}$ = 3.9 Hz), 93.4 (d, J$_{CF}$ = 175 Hz), 54.9 (d, J$_{CF}$ = 22.9 Hz), 46.5, 32.5 (d, J$_{CF}$ = 21.3 Hz), 18.2, 10.9; LC/MS (ESI-positive) m/z: 542.9 (M + H$^+$).

2-Tert-butyl-5-(2-chloropyrimidin-4-yl)-4-[3-[[ethyl (methyl)sulfamoyl]amino]-2-fluoro-phenyl]thiazole. To a solution of 3-[2-tert-butyl-5-(2-chloropyrimidin-4-yl) thiazol-4-yl]-2-fluoroaniline (102 mg, 0.281 mmol) in dichloromethane (1 ml) was added pyridine (0.5 ml) followed by ethyl(methyl)sulfamoyl chloride (265 mg, 1.68 mmol). The reaction was allowed to stir at 50 °C for 96 h. The reaction was worked up by extraction with ethyl acetate and 0.1 M HCl (aq). The product was purified by flash chromatography (5–30% ethyl acetate in hexanes) which gave impure material. This material was again purified by flash chromatography (0.5–6% methanol in dichloromethane). This provided the title compound (55 mg, 41% yield), which was used in the next step. ^{1}H NMR (400 MHz, CD$_3$CN) 8.45 (d, J = 5.4 Hz, 1 H), 7.66 (t, J = 7.5 Hz, 1 H), 7.55 (s, 1 H), 7.38 (t, J = 7.5 Hz, 1 H), 7.34 (dd, J = 8.0 Hz, 1 H), 7.04 (d, J = 5.4 Hz, 1 H), 3.19 (q, J = 7.2 Hz, 2 H), 2.79 (s, 3H), 1.51 (s, 9 H), 1.09 (t, J = 7.3 Hz, 3 H); LC/MS (ESI-positive) m/z: 484.2 (M + H$^+$).

5-(2-Aminopyrimidin-4-yl)-2-tert-butyl-4-[3-[[ethyl(methyl)sulfamoyl]amino]-2-fluoro-phenyl]thiazole (PLX7922). A solution of 2-tert-butyl-5-(2-chloropyrimidin-4-yl)-4-[3-[[ethyl(methyl)sulfamoyl]amino]-2-fluoro-phenyl]thiazole (51 mg, 0.11 mmol) dissolved in 5 ml of 7 M ammonia in methanol in a sealed reaction vial was placed in an oil bath at 80 °C and allowed to stir. After 48 h, the reaction was concentrated under reduced pressure and the resulting residue was purified by reverse-phase HPLC to provide the title compound, after lyophilization, as a white solid (31 mg, 61% yield). ^{1}H NMR (400 MHz, DMSO-d6) 9.71 (br s, 1 H), 8.04 (d, J = 5.1 Hz, 1 H), 7.54 (m, 1 H), 7.30 (m, 2 H), 6.77 (br s, 2 H), 6.03 (d, J = 5.1 Hz, 1 H), 3.06 (q, J = 7.0 Hz, 2 H), 2.67 (s, 3 H), 1.41 (s, 9 H), 0.99 (t, J = 7.0 Hz, 3 H); ^{13}C NMR (100 MHz, DMSO-d6) 181.9, 163.9, 159.3, 158.1, 152.2 (d, J$_{CF}$ = 251 Hz), 145.9, 134.7, 127.8, 126.9 (d, J$_{CF}$ = 13 Hz), 126.5, 125.2 (d, J$_{CF}$ = 5 Hz), 124.3 (d, J$_{CF}$ = 14 Hz), 105.8, 45.3, 38.1, 34.5, 30.8, 13.2; LC/MS (ESI-positive) m/z: 465.2 (M + H$^+$).

N-[3-[(5-chloro-1H-pyrrolo[2,3-b]pyridin-3-yl)-hydroxy-methyl]-2,4-difluoro-phenyl]-4-(trifluoromethyl)benzenesulfonamide. To a solution of N-(2,4-difluoro-3-formyl-phenyl)-4-(trifluoromethyl)benzenesulfonamide (83.4 g, 0.228 mol) and 5-chloro-1H-pyrrolo[2,3-b]pyridine (34.8 g, 0.228 mol) in anhydrous methanol (350 ml) was added potassium hydroxide (38.4 g, 0.684 mol). The reaction mixture was stirred at room temperature, under nitrogen, for 3 h and poured into water (1 l). The product was extracted with ethyl acetate (2 × 800 ml). The organic layers were combined, washed with brine (800 ml), dried, and concentrated under reduced pressure to yield a brown solid. This solid was suspended in acetonitrile (10 vol) overnight with stirring and then cooled in an ice bath for 3 h. The solids were isolated by filtration, washed with a minimum of cold acetonitrile, and dried to provide the title compound (56.8 g, 48% yield). ^{1}H NMR (400 MHz, DMSO-d6) 11.77 (s, 1 H), 10.38 (s, 1 H), 8.17 (d, J = 2.3 Hz, 1 H), 7.88 (s, 4 H), 7.75 (d, J = 2.3 Hz, 1 H), 7.18 (m, 1 H), 7.17 (m, 1 H), 7.05 (t, J = 9.0 Hz, 1 H), 6.20 (d, J = 4.9 Hz, 1 H), 6.02 (d, J = 4.9 Hz, 1 H); (LC/MS (ESI-positive) m/z: 518.0 (M + H$^+$).

N-[3-(5-chloro-1H-pyrrolo[2,3-b]pyridine-3-carbonyl)-2,4-difluoro-phenyl]-4-(trifluoromethyl)benzenesulfonamide (PLX5568). To a solution of N-[3-[(5-chloro-1H-pyrrolo[2,3-b]pyridin-3-yl)-hydroxy-methyl]-2,4-difluoro-phenyl]-4-(trifluoromethyl)benzenesulfonamide (100 g, 0.193 mol) in tetrahydrofuran (2.5 l) was added to Dess–Martin periodinane (99.2 g, 0.234 mol) under nitrogen. When the reaction was complete, the mixture was poured into 1 M sodium thiosulfate (700 ml) and saturated sodium bicarbonate solution (700 ml) and then extracted with ethyl acetate (2 × 700 ml). The organic layers were combined, washed with brine (800 ml), dried, and concentrated under reduced pressure to yield a brown solid. This residue was stirred in ethyl acetate (1 l) and silica (100 g) for 45 min and diluted with hexane (500 ml). The mixture was poured through a plug of silica and the product was eluted with 50:50 hexane:ethyl acetate. The fractions containing the product were combined and concentrated under reduced pressure to yield crude product as a yellow solid. Recrystallization of the crude product from ethanol provided the title compound as a pale yellow solid (71 g, 71% yield). ^{1}H NMR (400 MHz, DMSO-d6) 13.13 (s, 1 H), 10.51 (s, 1 H), 8.43 (s, 1 H), 8.38 (d, $J = 2.4$ Hz, 1 H), 8.18 (d, $J = 2.4$ Hz, 1 H), 7.93 (s, 4 H), 7.44 (m, 1 H), 7.28 (m, 1 H); ^{13}C NMR (100 MHz, DMSO-d6) 180.7, 157.0 (dd, $J_{CF} = 247$ Hz, $J_{CF} = 7.3$ Hz), 153.5 (dd, $J_{CF} = 251$ Hz, $J_{CF} = 8.4$ Hz), 148.1, 144.0, 143.9, 139.9, 133.1 (q, $J_{CF} = 32.3$ Hz), 130.4 (d, $J_{CF} = 9.2$ Hz), 128.7, 128.1, 126.9 (q, $J_{CF} = 3.8$ Hz), 126.3, 123.8 (q, $J_{CF} = 273$ Hz), 121.3 (dd, $J_{CF} = 13.8$ Hz, $J_{CF} = 3.8$ Hz), 118.6, 118.4 (dd, $J_{CF} = 25.4$ Hz, $J_{CF} = 22.6$ Hz), 115.3, 113.1 (dd, $J_{CF} = 23.1$ Hz, $J_{CF} = 3.8$ Hz); (LC/MS (ESI-positive) m/z: 516.1 (M + H$^+$).

In vitro and in vivo studies. *Biochemical assays and kinome selectivity profiling.* The *in vitro* RAF kinase activities were determined by measuring phosphorylation of a biotinylated substrate peptide as described previously[25]. PLX7904 was also tested against a panel of 287 kinases at concentrations of 1 μM in duplicate. Kinases inhibited by over 50% were followed up by IC$_{50}$ determination. The 287 kinases represent all major branches of the kinome phylogenetic tree. The inhibition screen of 287 kinases was performed under contract as complementary panels at Invitrogen (Life Technologies) SelectScreen profiling service, DiscoverX KINOMEScan service, and Reaction Biology Corporation Kinase HotSpot service.

Cell culture experiments. The B9 cell line was a gift from A. Balmain. The SK-MEL-239 and SK-MEL-239-C3 cell lines were provided by D. Solit and N. Rosen. The IPC-298 cell line was purchased from DSMZ. All other cell lines (A375, A431, COLO829, HCT116, and SKBR3) were purchased from ATCC. All cell lines were authenticated at the source by STR profiling and tested negative for mycoplasma contamination before use. Compounds dilutions were done in 100% DMSO and these titrations were diluted 500-fold in culture medium when added to cells, resulting in a final 0.2% DMSO concentration. Final compound concentrations are listed in text and figures.

Phospho-ERK AlphaScreen assay. To determine the effects of compound treatment on phosphorylation of ERK1/2, cells were plated in a 96-well plate and treated with an eight-point titration of compound for 1 h at 37 °C before lysis. To detect pERK, cell lysates were incubated with streptavidin-coated AlphaScreen donor beads, anti-mouse IgG AlphaScreen acceptor beads, a biotinylated anti-ERK1/2 rabbit antibody, and a mouse antibody that recognized ERK1/2 only when it was phosphorylated on Thr202 and Tyr204. The biotinylated ERK1/2 antibody bound both to the streptavidin-coated AlphaScreen donor beads and to ERK1/2 (regardless of its phosphorylation state), and the phospho-ERK1/2 antibody bound to the acceptor beads and to ERK1/2 that was phosphorylated at Thr202/Tyr204. An increase in ERK1/2 phosphorylation at Thr202/Tyr204 brought the donor and acceptor AlphaScreen beads into close proximity, generating a signal that could be quantified on an EnVision reader (Perkin Elmer). Inhibition of ERK phosphorylation resulted in a loss of signal compared with DMSO controls.

Phospho-ERK immunoblot analysis. Western blots were performed by standard techniques and analysed on an Odyssey Infrared Scanner (Li-COR Biosciences). The following antibodies were used: pERK1/2 (T202/Y204) and ERK1/2 (Cell Signaling).

Growth inhibition assay. Cells were plated into a 96-well plate at a density of 3,000 cells per well and allowed to adhere overnight. Compounds were dissolved in DMSO, diluted threefold to create an eight-point titration, and added to cells. After incubation for 72 h, cell viability was examined using CellTiter-Glo (Promega).

Anchorage-independent growth assay. Twenty-five thousand B9 cells were plated in each well of a six-well plate with a bottom layer of 1% and a top layer of 0.4% low melting agar (Sigma A4018) containing RPMI1640 medium with 10% FBS. For the RAF inhibitor study, B9 cells grown in soft agar were treated with vemurafenib, PLX4720 or PLX7904 at the indicated concentrations, or DMSO at 0.2% final concentration for 3 weeks. For the EGFR ligand study, B9 cells grown in soft agar were treated with AREG (R&D Systems 989-AR), TGF-α (R&D Systems 239-A),

or HB-EGF (R&D Systems 259-HE) at the indicated concentrations for 3 weeks. For the vemurafenib and erlotinib combination study, B9 cells grown in soft agar were treated with vemurafenib, erlotinib, or a combination of the two compounds at the indicated concentrations, or DMSO for 3 weeks. Anchorage-independent colonies ≥ 100 μm were scored using AxioVision Rel 4.8 software (Carl Zeiss).

ELISA for detecting EGFR ligands. Twenty thousand B9 cells were plated in each well of a 96-well plate and treated with DMSO control or compounds at the indicated concentrations for 48 h. Cell supernatants were collected and cells were lysed using 1 × cell lysis buffer (CST 9803). The amounts of AREG, TGF-α, and HB-EGF in cell supernatants or cell lysates were determined with the use of ELISA Development kits (R&D Systems DY989, DY239, and 259-HE-050N) according to the manufacturer's instructions.

EGFR signalling assay. B9 cells were treated with 1 μM or 5 μM vemurafenib or control vehicle for the indicated times in the absence of serum. Supernatants from treated B9 cells were then collected and added to newly plated, serum-starved (overnight) B9 cells for 10 min. Cells were washed with 1 × PBS twice, lysed, and subjected to western blot analysis. pEGFR Y1068, EGFR, pAkt S473, and Akt antibodies were purchased from Cell Signaling Technology.

RAF dimerization assays. BRAF–CRAF heterodimerization was characterized in cell lysates. Cells were plated on 96-well dishes and allowed to adhere overnight at 37 °C. Cells were treated with compound or DMSO for 1 h at 37 °C before lysis in RIPA buffer containing protease and phosphatase inhibitors. The lysates were transferred to ELISA plates coated with a monoclonal CRAF capture antibody, and incubated overnight at 4 °C. Further incubations with a polyclonal BRAF detection antibody and a horseradish-peroxidase-labelled secondary antibody were done at room temperature. After incubation with TMB substrate and sulfuric acid, the signal was analysed by measuring absorbance at 450 nm on a Tecan Safire plate reader.

Microarray gene expression analysis. B9 cells were plated in 1 μM vemurafenib, 1 μM PLX7904 or 0.2% DMSO vehicle control and incubated for 17 h. Cells were harvested, total RNA was isolated (RNeasy Mini Kit, Qiagen), and gene expression was measured using Affymetrix Mouse420_2 chips following the manufacturer's instructions. Vemurafenib response genes were identified by requiring the ratio between the treated and vehicle control samples be more than 1.9 (upregulated) or less than 0.54 (downregulated).

Tumour xenograft studies. All animal studies were conducted in accordance with the Institute for Laboratory Animal Research Guide for the Care and Use of Laboratory Animals and the US Department of Agriculture's Animal Welfare Act and approved by the institutional review board at testing facilities. Sample size (number of mice per group) was selected to provide at least 80% power to detect a two s.d. difference of mean tumour volume between two groups with two-sided type I error = 1%. The same formulation was used for both COLO205 and B9 xenograft studies. The powder of the test compound was dissolved in pure N-methyl-2-pyrrolidone. Diluent consisted of PEG400:TPGS:Poloxamer 407:water (40:5:5:50). Before gavage administration, fresh stock of N-methyl-2-pyrrolidone compound solution (or N-methyl-2-pyrrolidone for vehicle) was thoroughly mixed with the diluents to make a uniform suspension. Dosing volume was 5 μl g^{-1}. On the last day of the efficacy study, blood samples were collected at 0, 2, 4, and 8 h after last dosing, two animals per time point, for pharmacokinetic analysis. Animals were fed a standard rodent diet and water was supplied *ad libitum*. Tumour measurements were taken with an electronic microcalliper three times weekly. In addition, body weights were recorded at these times. Test facility investigators were blinded to the group allocation during the experiment.

COLO205 tumour cells were cultured in DMEM 10% FBS 1% penicillin/streptomycin supplemented with bovine insulin, at 37 °C. Balb/C nude mice, female, 6–8 weeks old, weighing approximately 18–22 g, were inoculated subcutaneously at the right flank with COLO205 tumour cells (5 × 10^6) in 0.1 ml of PBS mixed with matrigel (50:50) for tumour development. The treatment was started when mean tumour size reached approximately 100 mm^3, with eight mice in each treatment group randomized to balance the average weight and tumour size. B9 cells were expanded in DMEM 10% FBS 1% penicillin/streptomycin. Upon trypsinization the cells were washed three times with 20 ml RPMI, and after the final centrifugation were re-suspended, counted, and adjusted by volume to a final concentration of 5 × 10^7 cells per millilitre. B9 xenografts were started by injection of 5 × 10^6 cells subcutaneously in 6- to 7-week-old female nude Balb/c mice. Compound dosing started when the average size of tumours reached 50–70 mm^3. Animals were equally distributed over treatment groups ($n = 10$) to balance the average tumour size and body weight. Animals were dosed orally for days 1–14 twice daily and days 15–28 once daily with vehicle, vemurafenib 50 mg per kg, or PLX7904 50 mg per kg. 12-O-tetradecanoylphorbol-13-acetate (TPA) was put on the skin of all mice twice a week during weeks 3 and 4 at a dose of 2 μg in 200 μl acetone.

Crystallization and structure determination. Expression and purification of BRAF and BRAF[V600E] were performed as previously described[13,25]. Crystallization drops were prepared by mixing the protein solution with 1 mM of compound and the same amount of reservoir, and drops were incubated by vapour diffusion (sitting drops) at 4 °C. The mother liquor used to obtain co-crystals of PLX7904, dabrafenib, and PLX7922 with BRAF[V600E] consisted of 0.1 M BisTris at PH 6.0, 12.5% 2,5-hexane-diol, 12% PEG3350; the reservoir used to obtain co-crystals of PLX5568 with BRAF[WT] contained 0.1 M MES at pH 6.0, 35% (v/v) 2-methyl-2,4- pentanediol, and 0.2 M Li_2SO_4. All co-crystals were flash-frozen with liquid nitrogen, but BRAF[V600E] co-crystals were soaked in a solution containing the mother liquor plus 20% glycerol, before flash-freezing. X-ray diffraction data were collected at beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley Laboratory) and beamline 9.1 at Stanford Synchrotron Radiation Lightsource (Stanford University). Data were processed and scaled using MOSFLM[31] and SCALA in the CCP4 package[32]. All co-structures were solved using molecular replacement with the program MOLREP[33]. The starting models used were the inhibitor-bound BRAF[V600E] and BRAF[WT], respectively (Protein Data Bank accession numbers 4FK3 and 1UWJ). The final models were obtained after several rounds of manual rebuilding and refinement with PHENIX[34] and REFMAC[35]. A summary of the crystallography statistics is included in Extended Data Table 1.

31. Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. D* **55,** 1690–1695 (1999).
32. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67,** 235–242 (2011).
33. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66,** 22–25 (2010).
34. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
36. Joseph, E. W. *et al.* The RAF inhibitor PLX4032 inhibits ERK signalling and tumor cell proliferation in a V600E BRAF-selective manner. *Proc Natl Acad Sci USA* **107,** 14903–14908 (2010).
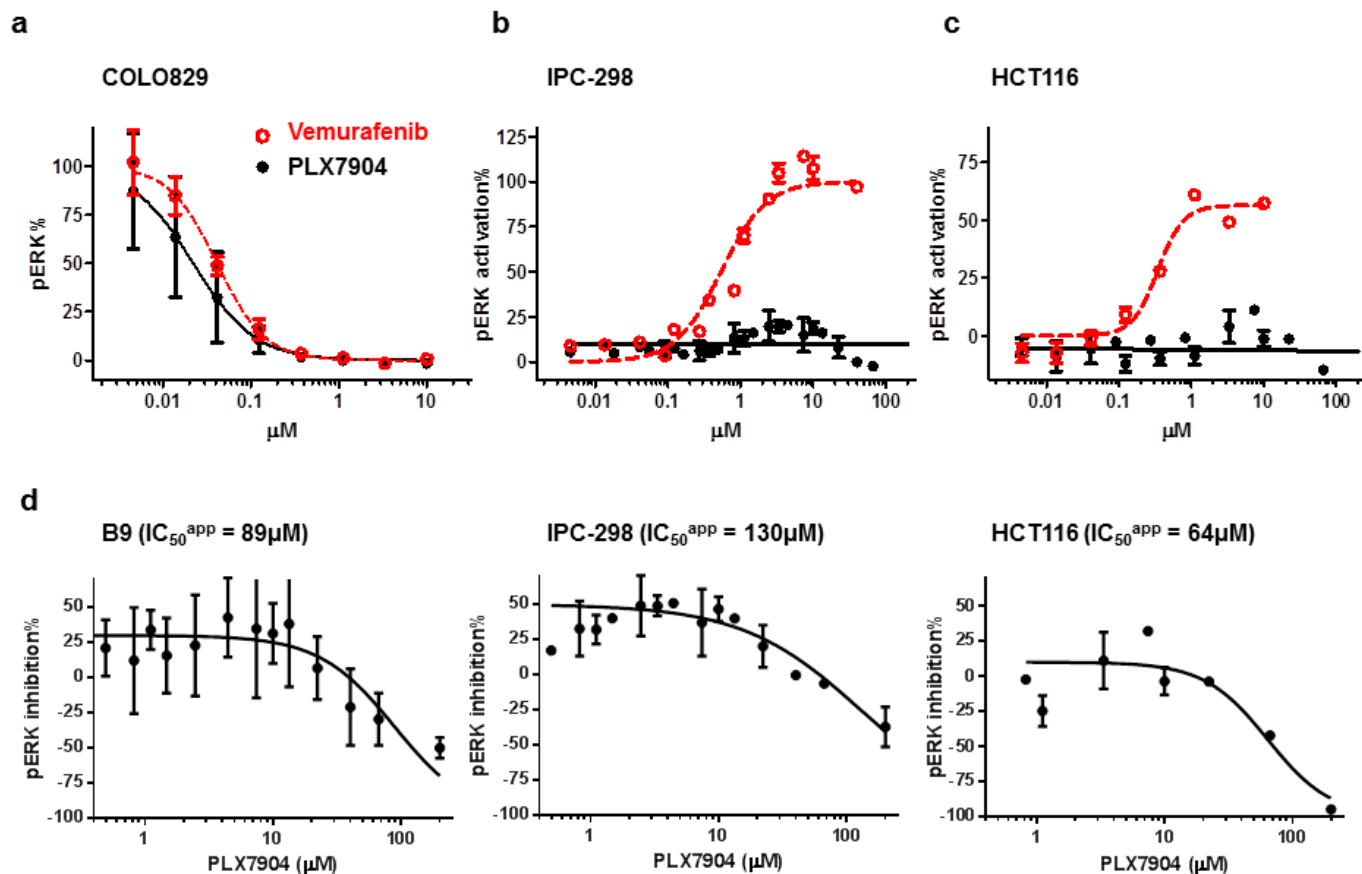
Extended Data Figure 1 | Differential effects of PLX7904 and vemurafenib on MAPK signalling. PLX7904 (black) and vemurafenib (red) show similar potency to block pERK signalling in human BRAF[V600E] melanoma cell COLO829 (**a**); but in RAS activated human melanoma cell line IPC-298 (NRAS[Q61L]) (**b**) and human colorectal carcinoma cell line HCT116 (KRAS[G13D]) (**c**), vemurafenib paradoxically activates MAPK signalling while PLX7904 causes negligible pERK increase. **d**, Expanded view of the pERK curves showing that PLX7904 inhibits pERK at high concentrations in three RAS mutant cell lines, with apparent IC_{50} (IC_{50}^{app}) values in the 100 μM range. Therefore, paradox breakers are not expected to affect the MAPK pathway in normal tissues (either paradoxical activation or inhibition) at therapeutic concentrations. The pERK curves were generated using an AlphaScreen assay. Mean ± s.d., $n = 5$ independent experiments.

a

Upregulated

Vem. | PLX7904
1 | 2 | 1 | 2

Areg
Hbegf
Egr1

Spry2

Tgfa
Fos

b

Downregulated

Vem. | PLX7904
1 | 2 | 1 | 2

Id2 —— | —— Cyp1b1

c
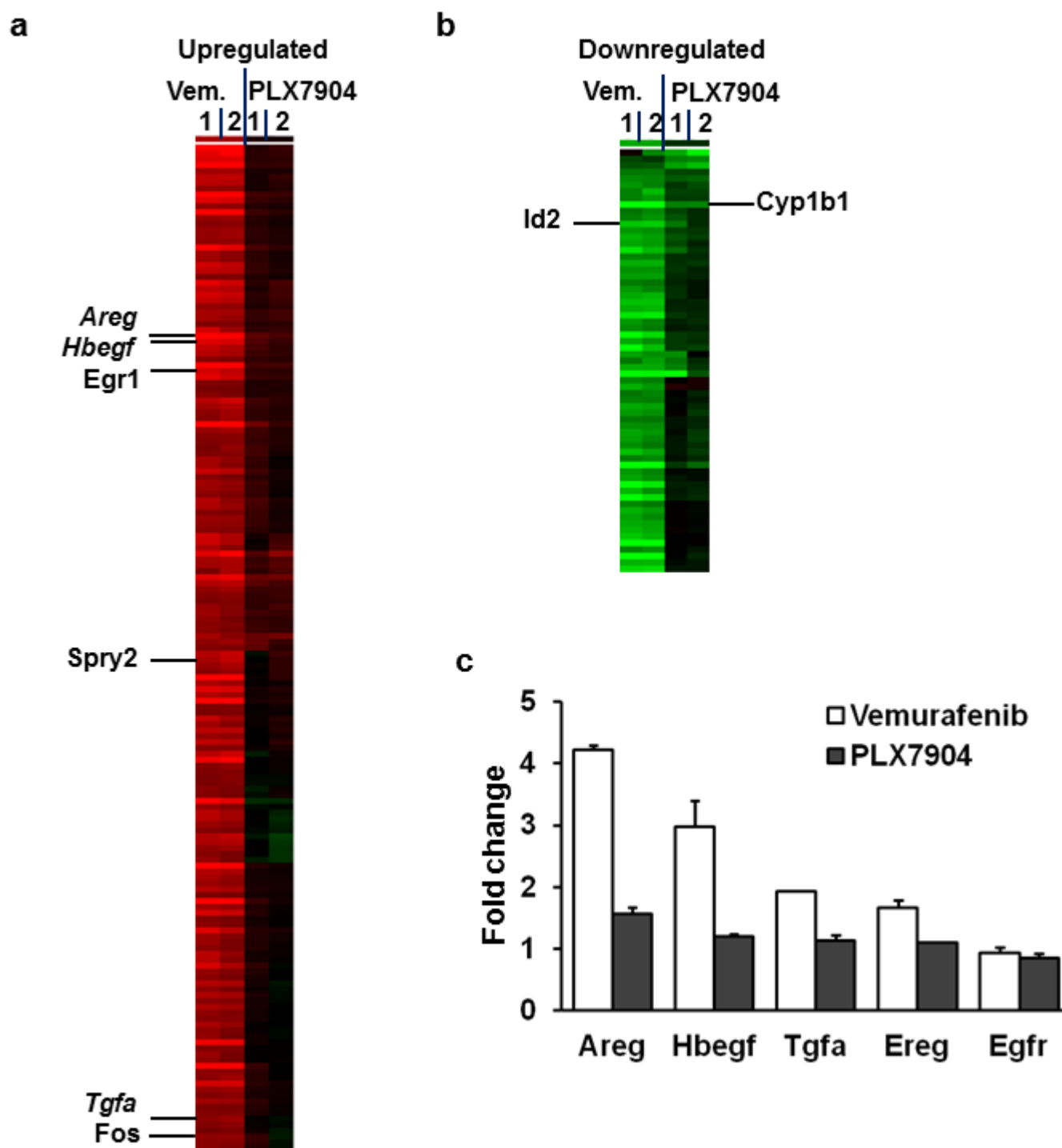
Extended Data Figure 2 | Gene expression analysis of B9 cells in response to either vemurafenib or PLX7904 treatment (both at 1 μM concentration). a, b, Hierarchical clustering of the 236 Affymetrix mouse gene probes (see Supplementary Table 2 for a complete list) that were upregulated (a) or downregulated (b) by vemurafenib (233 probes) or PLX7904 (4 probes). The single overlap, Cyp1b1, and four representative MAPK pathway-responsive genes as well as three genes that encode EGFR ligands are marked. Two independent experiments are shown. MAPK pathway response genes Spry2,

Fos, and Egr1 were upregulated by vemurafenib. The corresponding human genes are known to be suppressed by vemurafenib in BRAF[V600E] mutant human melanoma[36]. Opposing changes in expression were also observed with the Id2 gene. c, Changes in the messenger RNA levels of four EGFR ligands (amphiregulin, HB-EGF, TGF-α, and epiregulin) along with EGFR itself in B9 cells treated with vemurafenib or PLX7904. All four EGFR ligands abundantly expressed in B9 cells were induced by vemurafenib, but the expression of EGFR and other ERBB family members remained unchanged.
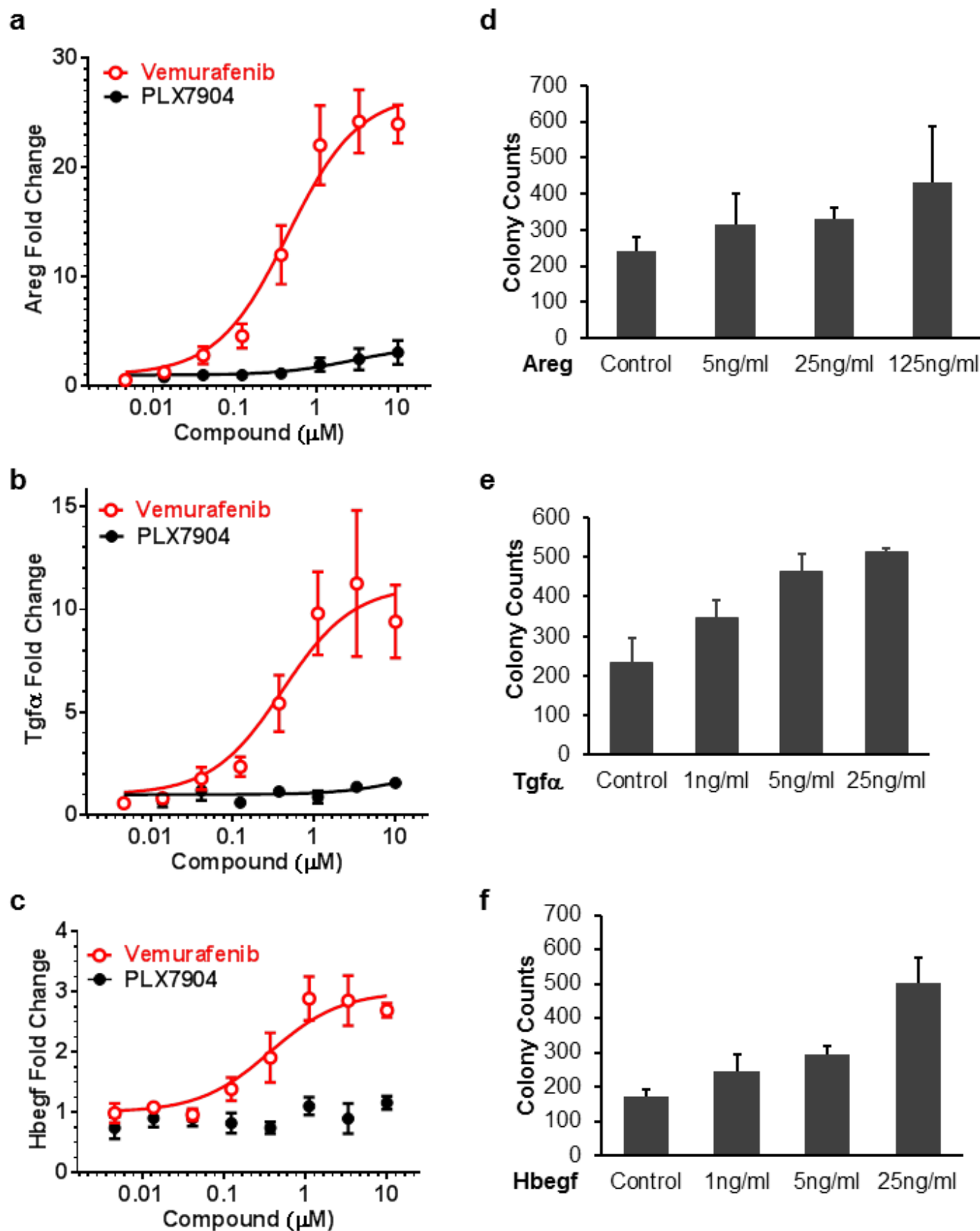
Extended Data Figure 3 | EGFR ligands may mediate vemurafenib-induced cuSCC. ELISA assays demonstrate increased levels of amphiregulin (**a**) and TGF-α (**b**) proteins in the supernatants and HB-EGF (**c**) in the cell lysates of B9 cells after vemurafenib treatment for 48 h. PLX7904 does not induce the expression of EGFR ligands. Like vemurafenib, exogenous amphiregulin (**d**), TGF-α (**e**), and HB-EGF (**f**) promote the anchorage-independent growth of B9 cells. B9 cells grown in soft agar were treated with EGFR ligands at the indicated concentrations for 3 weeks. Error bars, s.d. (**a–c**), s.e.m. (**d–f**); $n = 5$ (**a–c**) and 6 (**d–f**) independent experiments.
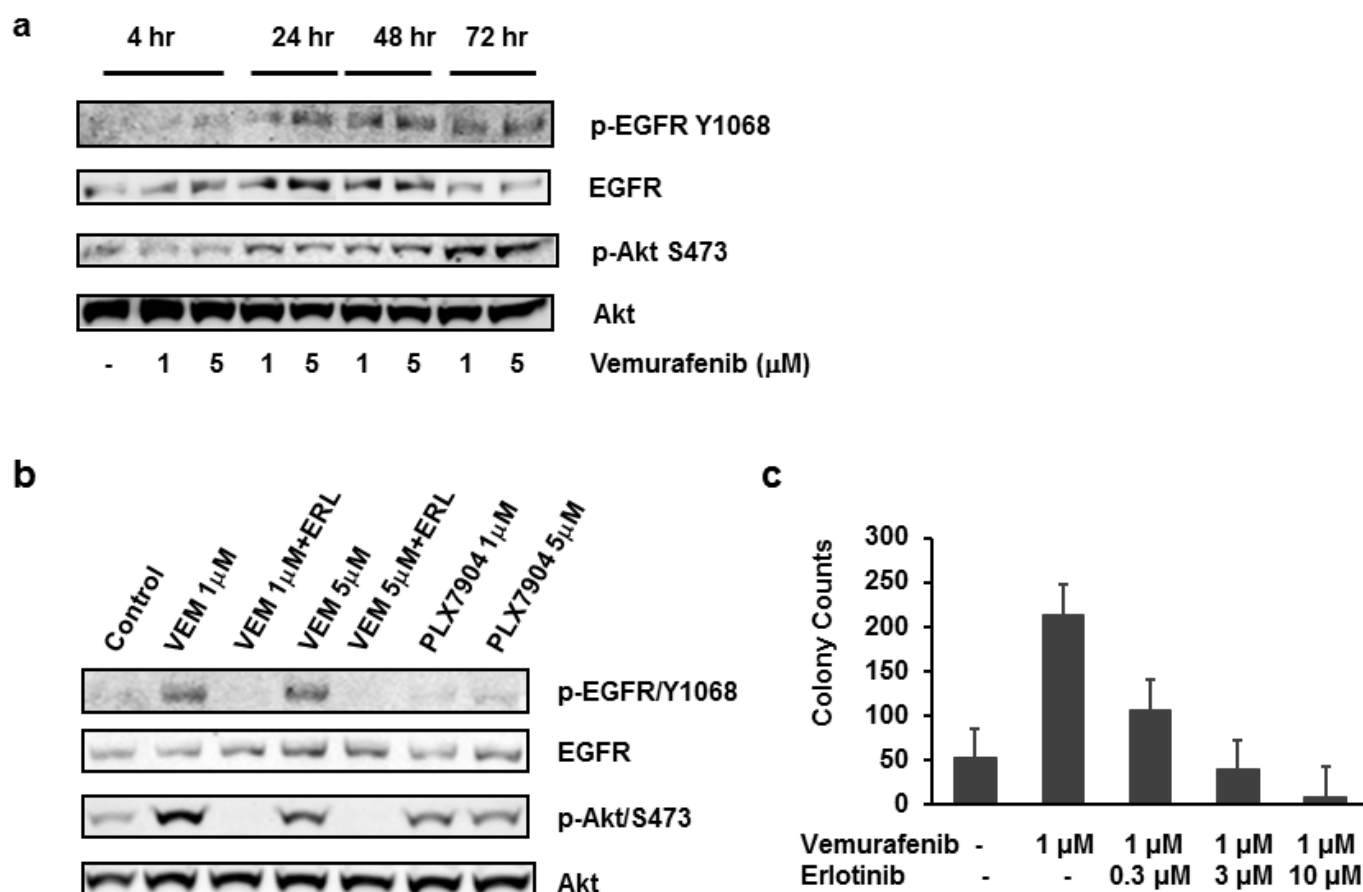
Extended Data Figure 4 | Effect of BRAF inhibitors on EGFR signalling.
a, EGFR signalling measured by levels of phosphorylated EGFR and AKT after a brief (10 min) exposure of serum-starved B9 cells to supernatant collected from B9 cells treated with vemurafenib for the indicated time. **b**, Pre-treatment with EGFR inhibitor erlotinib (ERL) inhibited EGFR signalling induced by supernatants from vemurafenib (VEM)-treated B9 cells. Serum-starved B9 cells were pre-treated with 3 μM erlotinib before starting a 10 min exposure to the

supernatants. Supernatants were collected from B9 cells treated with vemurafenib or PLX7904 for 3 days. **c**, Erlotinib inhibits the soft agar colony forming capacity of vemurafenib in B9 cells. Panels **a** and **b** are representative of results from three independent experiments. Error bars in **c**, s.e.m.; $n = 6$ independent experiments. Full scans of western blot data are presented in Supplementary Figure 1.
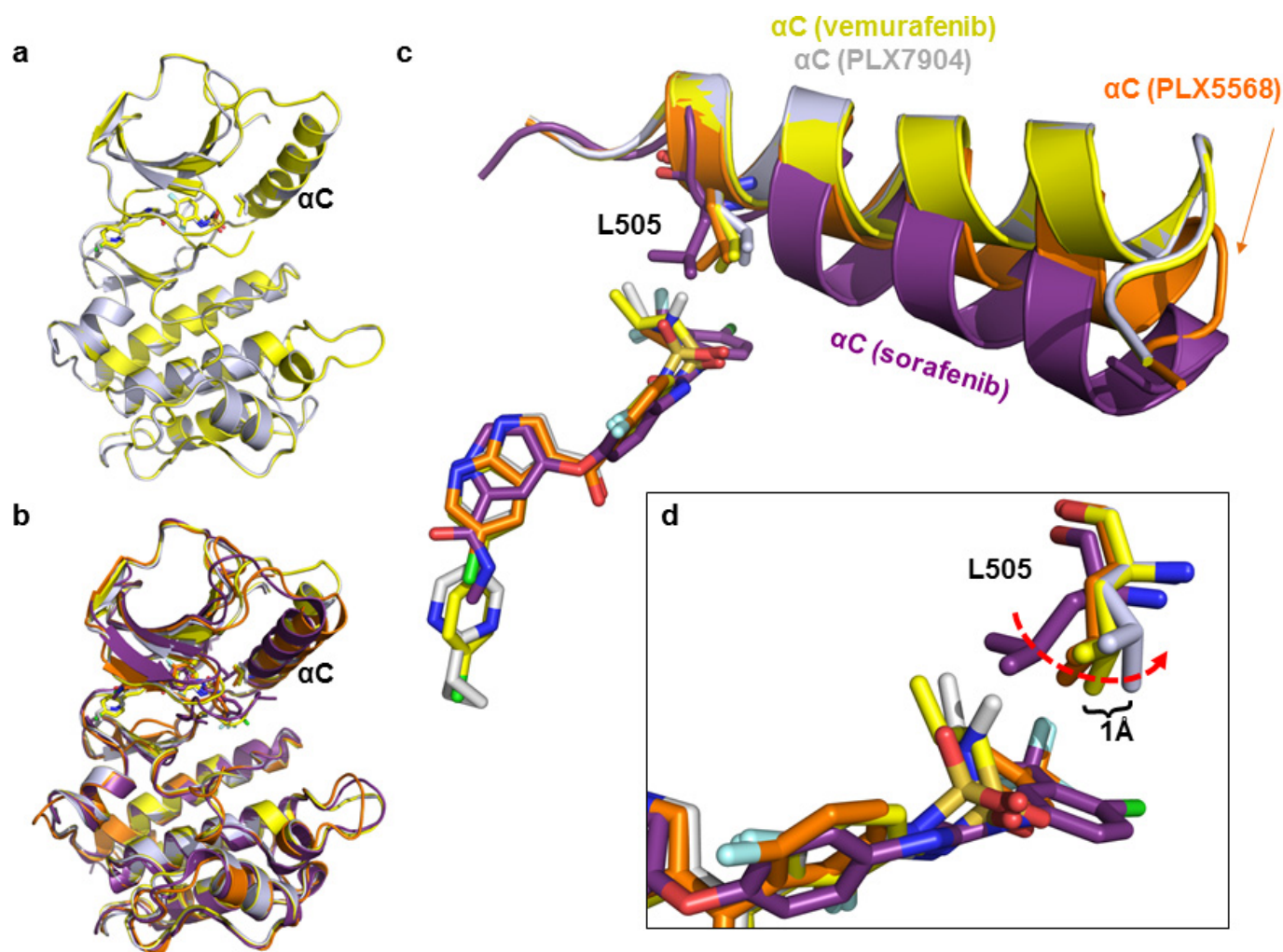
Extended Data Figure 5 | Comparison of inhibitor-bound BRAF structures.
a, Perfect alignment between vemurafenib and PLX7904-bound BRAF structures (backbone root mean squared deviation 0.22 Å). **b**, An overlay of the structures of BRAF bound to four inhibitors: sorafenib, PLX5568, vemurafenib, and PLX7904 (colour schedule same as **c**). **c**, Outward movement of αC helix in response to different inhibitors. From sorafenib to PLX5568 to vemurafenib, the degree of outward shift correlates with increasing ERK pathway inhibition index (Table 1). **d**, Close-up view showing the Leu505 side-chain position in the four structures. PLX7904 pushes the tip of Leu505 side-chain away by 1 Å from its position in the vemurafenib-bound structure.
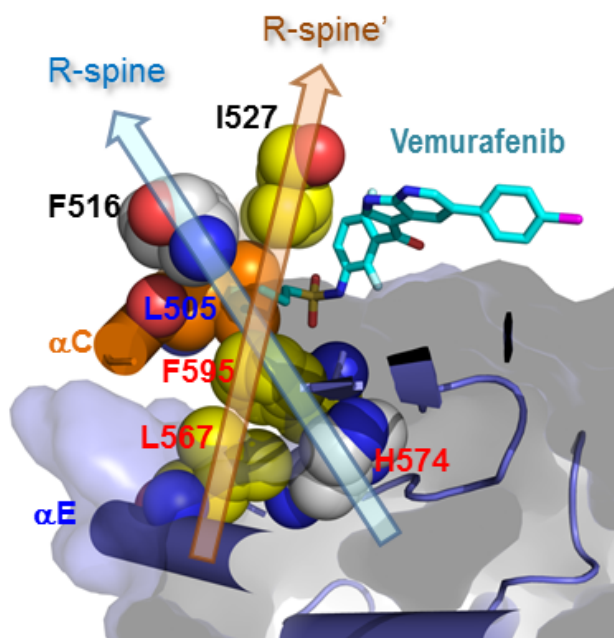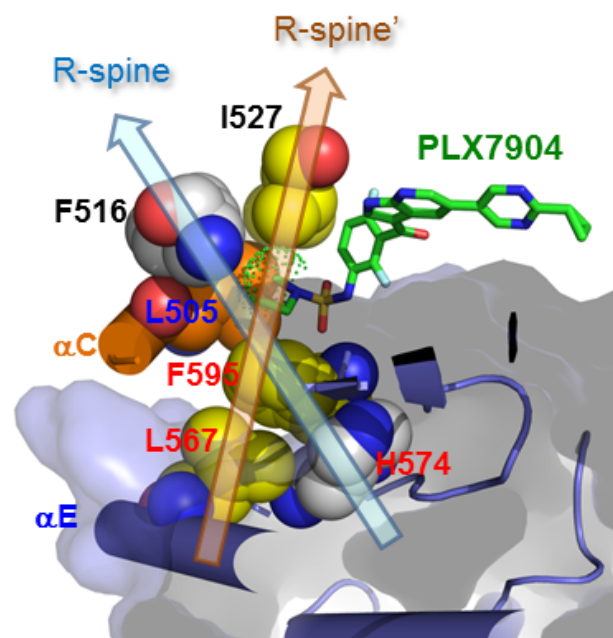
Extended Data Figure 6 | The regulatory spine (R-spine) in BRAF.
a, b, R-spine refers to four conserved hydrophobic residues that form a column in the active state of a kinase, and the distortion or disassembly of the spine marks the transition to an inactive state[22]. The term was introduced using PKA as the template, and the four residues that compose the R-spine included Leu95, Leu106, Tyr164, and Phe185 (PKA numbering). The corresponding residues in BRAF are Leu505, Phe516, His574, and Phe595 (rendered here in spheres). Tyr164 of PKA, which is a histidine (for example, His574 in BRAF) in most other kinases, forms hydrogen bonds with the backbone of the DFG motif and packs against the side chain of DFG Phe185 (the corresponding residue in

BRAF is Phe595). In the BRAF structure, Leu567 from αE helix also makes direct hydrophobic contacts with Phe595, an interaction that is conserved across the kinome. Leu567, Phe595, Leu505, along with another hydrophobic residue Ile527 that packs against Leu505, form a column (dubbed here as R-spine′) with an axis tilted 45° from that of the R-spine. Analyses of published kinase structures show that all four R-spine′ residues could be involved in kinase inhibitor binding whereas the two outer residues of R-spine rarely make direct contacts with inhibitors. Therefore, R-spine′ is more relevant for studying inhibitor-induced conformational change in kinases.
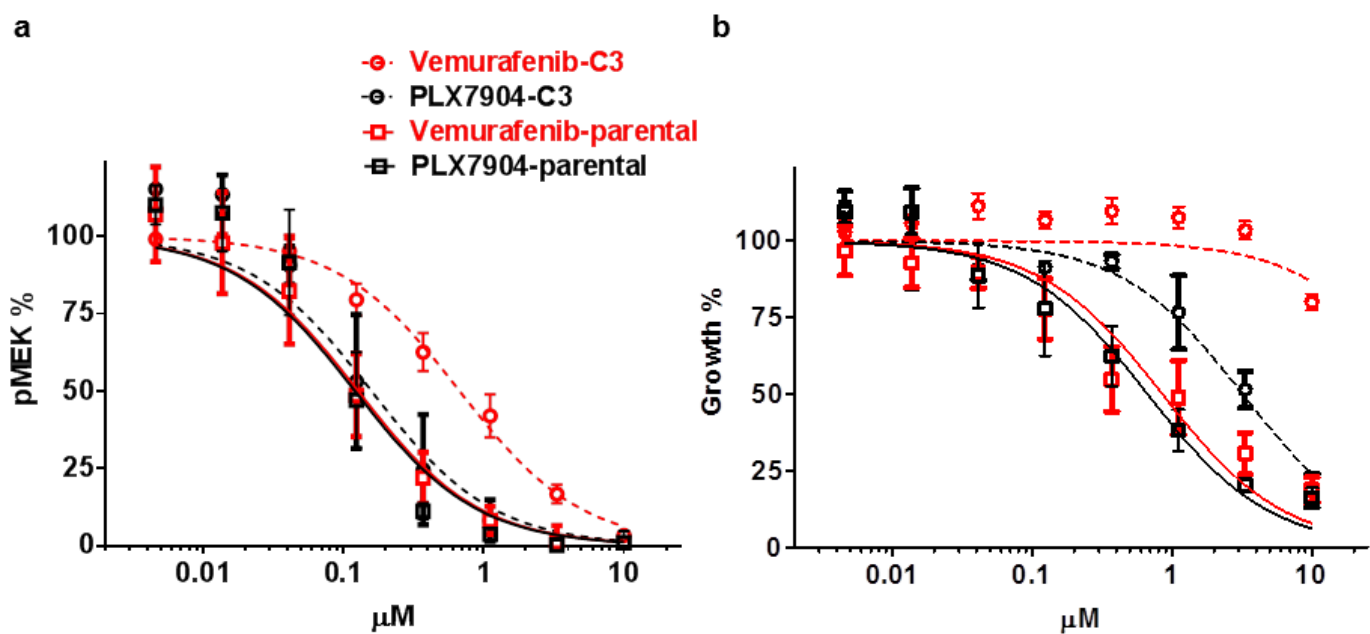
Extended Data Figure 7 | Vemurafenib-resistant cells remain relatively sensitive to paradox breakers. a, pMEK and **b**, growth IC_{50} curves (mean ± s.d., $n = 5$ experiments) for vemurafenib and PLX7904 in the SKMEL-239 parental cell line and a representative vemurafenib-resistant clone (C3) that expresses a spliced variant of $BRAF^{V600E}$ promoting dimerization.

Extended Data Table 1 | Data collection and refinement statistics

	$\mathrm{BRAF}^{\mathrm{V600E}}$-PLX7904 (PDB:4XV1)#	$\mathrm{BRAF}^{\mathrm{V600E}}$-Dabrafenib (PDB:4XV2)#	$\mathrm{BRAF}^{\mathrm{V600E}}$-PLX7922 (PDB:4XV3)#	$\mathrm{BRAF}^{\mathrm{WT}}$-PLX5568 (PDB:4XV9)#
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P4_322$
Cell dimensions				
$\quad a$, b, c (Å)	51.3, 104.8, 110.2	53.7, 105.7, 109.7	51.9, 105.4, 111.3	119.3, 119.3, 52.5
Resolution (Å)	76.0-2.47 (2.58-2.47) *	109.7-2.50 (2.64-2.50)	111.3-2.80 (2.95-2.80)	50.0-2.00 (2.05-2.00)
R_{sym} or R_{merge}	0.077 (0.743)	0.071 (0.531)	0.108 (0.591)	0.078 (0.354)
$I/\sigma I$	7.8 (1.0)	8.1 (1.4)	5.1 (1.5)	6.9 (2.1)
Completeness (%)	100.0 (100.0)	99.9 (99.9)	99.9 (100.0)	99.9 (100.0)
Redundancy	5.8 (5.9)	6.1 (6.2)	4.6 (4.8)	13.6 (11.2)
Refinement				
Resolution (Å)	76.0-2.47	54.9-2.50	55.7-2.80	50.0-2.00
No. reflections	22,006	22,263	15,592	24,901
$R_{\mathrm{work}}/ R_{\mathrm{free}}$	0.234/0.273	0.212/0.244	0.258/0.296	0.204/0.238
No. atoms				
$\quad$ Protein	3912	3970	3864	2079
$\quad$ Ligand/ion	36	70	31	34
$\quad$ Water	40	54	44	256
B-factors				
$\quad$ Protein	73	63	62	27
$\quad$ Ligand/ion	89	41	68	22
$\quad$ Water	56	48	35	35
R.m.s deviations				
$\quad$ Bond lengths (Å)	0.003	0.003	0.003	0.006
$\quad$ Bond angles (°)	0.7	0.7	0.7	1.1

*Highest resolution shell is shown in parenthesis.
Data for each structure is collected from a single crystal.

LETTER

Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria

Gunter Wegener[1,2]*, Viola Krukenberg[1]*, Dietmar Riedel[3], Halina E. Tegetmeyer[4,5] & Antje Boetius[1,2,4]

The anaerobic oxidation of methane (AOM) with sulfate controls the emission of the greenhouse gas methane from the ocean floor[1,2]. In marine sediments, AOM is performed by dual-species consortia of anaerobic methanotrophic archaea (ANME) and sulfate-reducing bacteria (SRB) inhabiting the methane–sulfate transition zone[3–5]. The biochemical pathways and biological adaptations enabling this globally relevant process are not fully understood. Here we study the syntrophic interaction in thermophilic AOM (TAOM) between ANME-1 archaea and their consortium partner SRB HotSeep-1 (ref. 6) at 60 °C to test the hypothesis of a direct interspecies exchange of electrons[7,8]. The activity of TAOM consortia was compared to the first ANME-free culture of an AOM partner bacterium that grows using hydrogen as the sole electron donor. The thermophilic ANME-1 do not produce sufficient hydrogen to sustain the observed growth of the HotSeep-1 partner. Enhancing the growth of the HotSeep-1 partner by hydrogen addition represses methane oxidation and the metabolic activity of ANME-1. Further supporting the hypothesis of direct electron transfer between the partners, we observe that under TAOM conditions, both ANME and the HotSeep-1 bacteria overexpress genes for extracellular cytochrome production and form cell-to-cell connections that resemble the nanowire structures responsible for interspecies electron transfer between syntrophic consortia of *Geobacter*[9,10]. HotSeep-1 highly expresses genes for pili production only during consortial growth using methane, and the nanowire-like structures are absent in HotSeep-1 cells isolated with hydrogen. These observations suggest that direct electron transfer is a principal mechanism in TAOM, which may also explain the enigmatic functioning and specificity of other methanotrophic ANME–SRB consortia.

The anaerobic oxidation of methane with sulfate (AOM) controls the emission of methane from the seabed[1,3,4]. At environmental conditions the net reaction $CH_4(aq) + SO_4^{2-} \rightarrow HS^- + HCO_3^- + H_2O$ allows an energy yield of only −20 to −40 kJ per mol of methane oxidized, shared between the two partner organisms. Generally, AOM consortia show exceptionally slow growth with generation times >2 months, which has so far impeded their cultivation[6,11]. Sulfate-coupled AOM in marine habitats is performed by members of three different ANME clades (ANME-1, -2 and -3), which associate physically with specific partner bacteria of the *Desulfosarcina/Desulfococcus* or the *Desulfobulbus* cluster[11–13], indicating an obligate functional role of the SRB in AOM. Early studies had already suggested a syntrophic coupling of both partners via a transfer of reducing equivalents[4,14], yet the underlying mechanisms remain unknown. Biochemically, the anaerobic oxidation of methane appears in the ANME and involves a reversal of the enzymatic machinery of the methanogenesis pathway[15–18]. However, reversing an energy-yielding process is per se endergonic, and hence AOM requires an efficient transfer of reducing equivalents from methane to sulfate, so that the ANME can gain energy by AOM[14,19]. Previous results indicate that the partner SRB[4,20] act as electron sinks of AOM, but recently members of the ANME-2 clade were also suggested to perform incomplete sulfate reduction by an as yet unknown pathway[21].

In this study we focus on the hypothesis of syntrophic growth in thermophilic AOM consortia by direct interspecies electron transfer, and test this and alternative hypotheses (mechanisms illustrated in Extended Data Fig. 1). The studied sediment-free TAOM enrichment was cultivated at 60 °C and supplied with 28 mM sulfate and 0.2 MPa methane, allowing an energy yield (ΔG_R) of −34 kJ mol^{-1}, and resulting in doubling times of approximately 68 days (Fig. 1a) and growth efficiencies of 2% (see Methods). The culture is dominated by consortia of ANME-1 and HotSeep-1 appearing in an approximate 1:1 stoichiometry. Owing to their larger size ANME account for around 75% of the consortial biomass (Fig. 1b and Extended Data Table 1). Using a dilution-to-extinction approach (1:10 to 1:10^9) with hydrogen (0.2 MPa) and sulfate (28 mM), we were able to separate a strain of HotSeep-1 that was identical to the partner bacterium of the
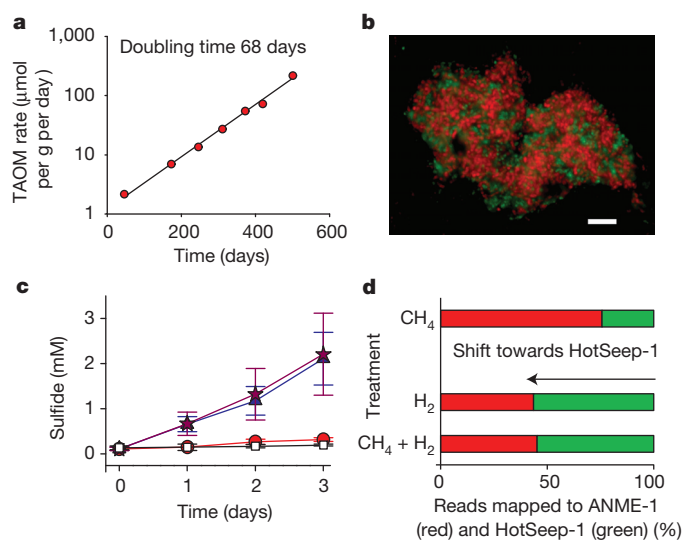
Figure 1 | Activity of the TAOM consortia in culture. a, The exponential increase of sulfide production translates to a doubling time of 68 days (biological replicates $n = 4$). **b**, Representative fluorescence micrograph of TAOM consortia from **a**; red, ANME-1; green, HotSeep-1. Scale bar, 10 μm. Representative of 14 similar images recorded. **c**, Sulfide production under TAOM conditions (red circles, 0.07 mM sulfide per day) versus a control (white squares, 0.02 mM per day). Hydrogen (blue triangles), or hydrogen plus methane (purple stars) increased sulfide production (both 0.55 mM per day; biological replicates $n = 3$, symbols represent mean values, error bars are s.d.). **d**, Per cent of total RNA reads mapped to ANME-1 (red) and HotSeep-1 (green) after 3 days of incubation (**c**), biological replicates $n = 3$, data are mean values, for statistical analyses see Supplementary Table 1.

[1]Max-Planck Institute for Marine Microbiology, 28359 Bremen, Germany. [2]MARUM, Center for Marine Environmental Sciences, University Bremen, 28359 Bremen, Germany. [3]Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [4]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, 27570 Bremerhaven, Germany. [5]Center for Biotechnology, Bielefeld University, 33615 Bielefeld, Germany.
*These authors contributed equally to this work.

TAOM consortium (>99% identity in 16S ribosomal RNA gene and internal transcribed spacer region, and (genomic) average nucleotide identity >99%, Extended Data Table 2). This strain grows without ANME-1 as single cells or in mono-species aggregates (Extended Data Fig. 2a) and with a single contaminant, *Archaeoglobus* sp. (1–5% of all cells), which does not form consortia with HotSeep-1. Substrate tests with the HotSeep-1 culture showed that it is an obligate chemo-lithoautotroph, with hydrogen and sulfate as the sole molecular redox couple and doubling times of 4 to 6 days (see Methods and Extended Data Fig. 2b, c). Although the supplied hydrogen (0.2 MPa) could provide a tenfold higher energy yield to HotSeep-1 than syntrophy in TAOM ($\Delta G_R = -151 \text{ kJ mol}^{-1}$ versus -17 kJ mol^{-1}, the latter being half of the net energy yield of TAOM consortia), its carbon assimilation efficiency remained similarly low (approximately 1.5% of converted reducing equivalents).

We compared the activity of the TAOM consortia and the hydro-genotrophic HotSeep-1 by physiological experiments combined with metagenomic and metatranscriptomic analyses and electron microscopy of the involved organisms. A classical experiment for the study of syntrophy in dual-species consortia is the addition of potential intermediates that could be theoretically produced by the ANME as a by-product of methane oxidation, and consumed as the electron donor by the partner SRB (mechanisms illustrated in Extended Data Fig. 1). If these compounds were relevant in interspecies transfer of reducing equivalents, their addition to the medium should favour the electron-accepting partner, and should repress the electron transfer between the consortial partners[14,21–23]. By contrast, a model of direct electron transfer via nanowires as proposed in refs 7,8,10, would be insensitive to such external additions of potential intermediates if the additions do not represent an alternative, preferred substrate for one of the partner organisms.

With the exception of hydrogen, none of the potential intermediates added as sole electron donor caused significant microbial sulfide production in the TAOM enrichment (Extended Data Table 3). Carbon monoxide and methyl sulfide even inhibited sulfide production when added together with methane. Carbon monoxide is known to inhibit cytochrome *c* activity, which may play an important role in intra- and intercellular transfer of reducing equivalents in AOM[17]. Methylated substrates may interfere with the reverse, oxidative operation of the methanogenesis pathway[24]. The addition of colloidal zero-valent sulfur to the TAOM culture (supplied in concentrations from 1 to 25 mM, Extended Data Fig. 3a) did not result in the production of sulfide and sulfate as reported in a previous study with ANME2a/DSS consortia[21]. However, with hydrogen as an electron donor (0.16 MPa), sulfide production rates increased three- to eightfold compared to replicate incubations with methane as the sole electron donor at TAOM conditions (Fig. 1c and Extended Data Table 3). We investigated further the influence of hydrogen on the oxidation of methane using head-space-free incubations (Extended Data Fig. 4). In incubations with methane and hydrogen, hydrogen was first selectively consumed and methane oxidation was repressed. When hydrogen was consumed, methane oxidation rates recovered to the same level as in replicate incubations with only methane, suggesting an inhibition of methane oxidation in the presence of hydrogen. To investigate the influence of hydrogen on the consortial partners, we mapped total RNA reads to the genome drafts of ANME-1 and HotSeep-1 after exposure to different substrate conditions (Fig. 1d, for read numbers see Supplementary Table 1). Under TAOM conditions, relative RNA expression patterns reflected the biomass ratio between the ANME and their partner bacteria (3:1) (Fig. 1d). The addition of hydrogen caused a strong relative increase of HotSeep-1 over ANME gene expression, even in the presence of methane. This indicates that if the partner SRB does not act as an electron sink for reverse methanogenesis, ANME activity is repressed; an effect of syntrophic cooperation that was predicted previously[14].

To test the hypothesis of hydrogen production by ANME as a direct intermediate in TAOM (Extended Data Fig. 1) that is consumed by HotSeep-1, we assessed the presence and production of hydrogen under TAOM conditions. Maximal hydrogen concentrations were only about 2 Pa in the TAOM enrichments, and re-established within 7 h after gas phase exchange (Fig. 2a). Thermodynamically, HotSeep-1 could thrive on these low hydrogen concentrations with an energy yield of approximately -24 kJ mol^{-1}. However, the production of hydrogen in TAOM cultures corresponded to only ~0.5% of the theoretical hydrogen production rates as reflected by sulfide production (according to the stoichiometry of reverse methanogenesis; Fig. 2b). This is insufficient to explain the consortial growth of HotSeep-1. Furthermore, we could not detect catalytic subunits of [FeFe] or [NiFe] hydrogenases in the ANME-1 draft genome. In conclusion, hydrogen appears to be an alternative growth substrate for HotSeep-1 when available externally, but is not provided by ANME-1 as an intermediate in TAOM.

An alternative explanation of the TAOM interaction could be direct interspecies electron transfer (DIET) between ANME-1 and HotSeep-1, also hypothesized as a principle mechanism for syntrophic growth of AOM consortia[8,10,17,25]. A switch from interspecies hydrogen transfer to DIET has been previously shown for the dual-species interaction between *Geobacter sulfurreducens* and *Geobacter metallireducens*, benefiting both consortial partners, as evidenced by their increased growth rates via DIET[10,26]. In the tightly packed *Geobacter* consortia, a dense network of cell-to-cell connections was detected by transmission electron microscopy and immunogold labelling, probably serving in electron transfer[10]. The functioning of electron transfer via conductive cell-to-cell connections (nanowires) is not fully understood, but apparently involves the expression of the pilin protein PilA of the type IV pili together with certain members of the cytochrome *c* family[9,10,27,28]. Recent findings of such cytochromes in the ANME genome, along with redox-dependent staining of the intercellular matrix of the ANME-2/SRB consortia[8], suggest that DIET could also be relevant in AOM.

To find evidence for DIET in TAOM we analysed the genome and specific gene expression of ANME-1 and HotSeep-1, with the focus on similarities to the *Geobacter* consortia using DIET as the main electron transfer mechanism. The ANME-1 draft genome contains several potentially extracellular multi-haem cytochrome *c* proteins, some of which are highly expressed during TAOM, but no genes for pili formation (Extended Data Table 4, Supplementary Table 2). However, HotSeep-1 comprises the genes for the biosynthesis and assembly of
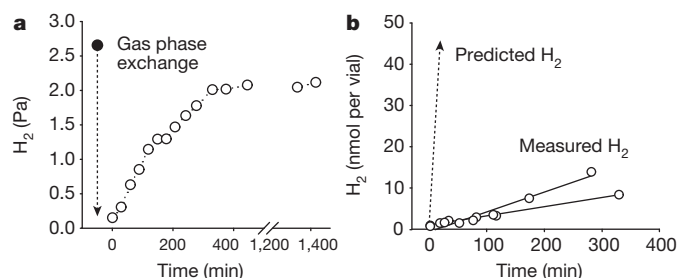
Figure 2 | Hydrogen in TAOM cultures. a, Hydrogen gas pressure under TAOM conditions (methane, 0.2 MPa; sulfate, 28 mM; 5-day incubation (filled circle)). The dashed arrow depicts gas phase exchange with methane. Open circles show equilibration of hydrogen in the headspace ($n = 1$). **b**, Hydrogen production in 10 ml TAOM culture supplied with 0.2 MPa methane after headspace exchange and addition of 10 mM molybdate (final concentration) to inhibit hydrogen consumption. Open circles are replicate measurements with hydrogen production of 2 and 3 nmol l^{-1} min^{-1}. Dotted line is predicted hydrogen production for reverse methanogenesis ($CH_4 + 2H_2O \rightarrow CO_2 + 4H_2$) = 420 nmol H_2 l^{-1} min^{-1} culture, for an observed sulfide production rate of 104 nmol l^{-1} min^{-1}. Both experiments were repeated once with the same results.

a mRNA expression relative to TAOM condition

0.00 0.01 0.10 1.00 10.00

ANME-1
- *mcrA* *
- *cytC* (A) *

HotSeep-1
- *dsrA* *
- *cytC* (H) *
- *pilA* *

Figure 3 | Expression of genes and visualization of structures attributed to interspecies electron transfer in TAOM. a, Expression of archaeal (red) and bacterial (green) genes in incubations with hydrogen (darker shade), or hydrogen plus methane (lighter shade), relative to methane alone (expression under TAOM condition = 1; biological replicates $n = 3$; *$P < 0.05$; for statistics see Supplementary Table 5b). **b–f**, Micrographs of thin sections. Scale bars, 300 nm. **b–d**, TAOM consortia with HotSeep-1 cells (H; rod-shaped; approximately $1 \times 0.5\,\mu m$) and ANME-1 cells (A; cylindrical shape with envelope[30]; $1.5 \times 0.8\,\mu m$). Nanowires of 10 nm diameter and up to several thousand nanometres in length connect both species, representative of 70 recorded images. **c, d**, Arrows mark the apparent origin of the wires from the membrane of HotSeep-1 bacteria to the polar sites of ANME-1 cells. **d**, HotSeep-1 cell with nanowires crossing the cell membrane (marked by the arrow). **e, f**, Aggregated HotSeep-1 cells grown with hydrogen do not develop nanowires; representative of 48 images.

type IV pili, as well as large multi-haem cytochrome *c* proteins, both with high amino acid similarity to respective proteins in *Geobacter* spp.[29] (Extended Data Table 4 and Supplementary Tables 3 and 4). We further investigated expression patterns of these potentially DIET-related genes in comparison to genes for AOM (*mcrA*) and sulfate reduction (*dsrA*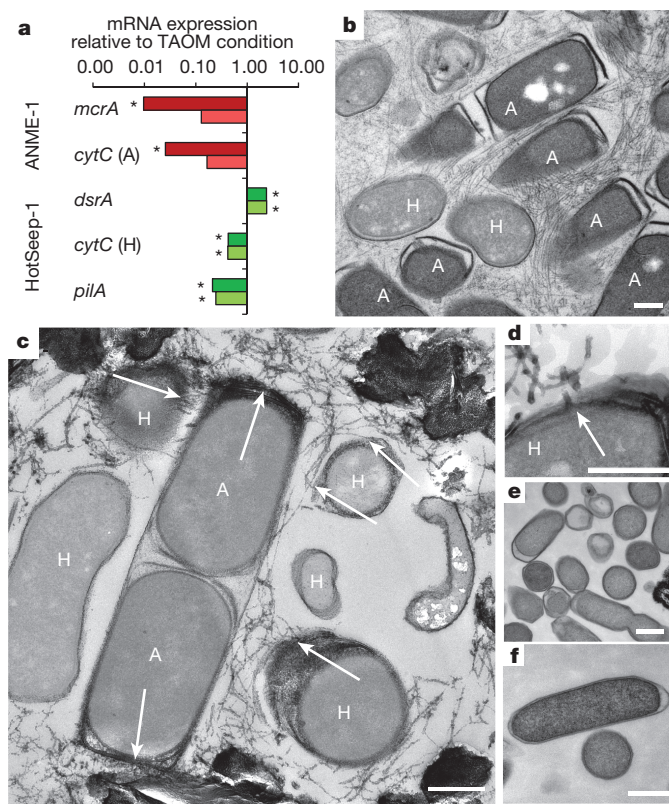), representing key catabolic processes in ANME-1 and HotSeep-1 (Fig. 3a, for statistical analyses see Supplementary Table 5). In agreement with the results from total RNA expression (Fig. 1d), a switch from methane to hydrogen (or methane plus hydrogen) as an energy source caused an immediate drop in *mcrA* and cytochrome expression in ANME, as well as a reduction of the expression of the HotSeep-1 pili and cytochromes. Comparing relative gene expression of HotSeep-1 in consortial growth using methane, versus single growth using hydrogen, both *pilA* and cytochrome *c* are clearly overexpressed under TAOM conditions, this is also the case when compared to *dsrA* expression (Extended Data Fig. 5, for statistical analyses see Supplementary Table 5).

This observation is supported independently by transmission electron microscopy on thin sections of TAOM consortia. Using two different embedding techniques we found a dense network of pili-like structures connecting HotSeep-1 to ANME-1 cells (Fig. 3b, c and Extended Data Fig. 6a), resembling the nanowire structures found in

Geobacter consortia (visualized with the same methods, see Extended Data Fig. 6b). In TAOM consortia the nanowires are larger, and appear more dense, with diameters of approximately 10 nm and apparent lengths of 100 to >1,000 nm. In agreement with the genomic patterns, these wires seem to be formed by the partner bacteria, connecting to the ANME-1 at their polar sides (Fig. 3b–d). In contrast, HotSeep-1 cells in mono-species aggregates isolated with hydrogen show smooth surfaces without such extracellular extensions (Fig. 3e, f), indicating that the observed intercellular structures are specific to consortial growth under TAOM conditions and not only related to cellular attachment.

In conclusion, our data show that consortial growth of thermophilic ANME-1 archaea and HotSeep-1 bacteria is probably based on similar principles as those in *Geobacter* consortia, where DIET is mediated by intercellular wiring made up of pili-like structures and outer-membrane multi-haem cytochromes. The underlying biophysics and biochemistry of intercellular wiring for direct electron transfer needs further investigation. If this mode of syntrophic cooperation between the electron-generating archaea and nanowire-producing bacteria is also the underlying mechanism for other types of AOM consortia as suggested recently[8], it may explain the enigmatic specificity of dual-species partnerships in AOM in general.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Reeburgh, W. S. Oceanic methane biogeochemistry. *Chem. Rev.* **107,** 486–513 (2007).
2. Boetius, A. & Wenzhöfer, F. Seafloor oxygen consumption fuelled by methane from cold seeps. *Nat. Geosci.* **6,** 725–734 (2013).
3. Hinrichs, K.-U., Hayes, J. M., Sylva, S. P., Brewer, P. G. & DeLong, E. F. Methane-consuming archaebacteria in marine sediments. *Nature* **398,** 802–805 (1999).
4. Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407,** 623–626 (2000).
5. Orphan, V. J. *et al.* Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Appl. Environ. Microbiol.* **67,** 1922–1934 (2001).
6. Holler, T. *et al.* Thermophilic anaerobic oxidation of methane by marine microbial consortia. *ISME J.* **5,** 1946–1956 (2011).
7. Thauer, R. K. & Shima, S. Methane as fuel for anaerobic microorganisms. *Ann. NY Acad. Sci.* **1125,** 158–170 (2008).
8. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* http://dx.doi.org/10.1038/nature15512 (this issue).
9. Reguera, G. *et al.* Extracellular electron transfer via microbial nanowires. *Nature* **435,** 1098–1101 (2005).
10. Summers, Z. M. *et al.* Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. *Science* **330,** 1413–1415 (2010).
11. Knittel, K. & Boetius, A. Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* **63,** 311–334 (2009).
12. Niemann, H. *et al.* Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* **443,** 854–858 (2006).
13. Schreiber, L., Holler, T., Knittel, K., Meyerdierks, A. & Amann, R. Identification of the dominant sulfate-reducing bacterial partner of anaerobic methanotrophs of the ANME-2 clade. *Environ. Microbiol.* **12,** 2327–2340 (2010).
14. Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. Field and laboratory studies of methane oxidation in an anoxic marine sediment: evidence for a methanogen-sulfate reducer consortium. *Glob. Biogeochem. Cycles* **8,** 451–463 (1994).
15. Krüger, M. *et al.* A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **426,** 878–881 (2003).
16. Hallam, S. J. *et al.* Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305,** 1457–1462 (2004).
17. Meyerdierks, A. *et al.* Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ. Microbiol.* **12,** 422–439 (2010).
18. Kojima, H., Moll, J., Kahnt, J., Fukui, M. & Shima, S. A reversed genetic approach reveals the coenzyme specificity and other catalytic properties of three enzymes putatively involved in anaerobic oxidation of methane with sulfate. *Environ. Microbiol.* **16,** 3431–3442 (2014).
19. Thauer, R. K. Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO_2. *Curr. Opin. Microbiol.* **14,** 292–299 (2011).
20. Milucka, J., Widdel, F. & Shima, S. Immunological detection of enzymes for sulfate reduction in anaerobic methane-oxidizing consortia. *Environ. Microbiol.* **15,** 1561–1571 (2013).
21. Milucka, J. *et al.* Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491,** 541–546 (2012).

22. Schink, B. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61,** 262–280 (1997).
23. Stams, A. J. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Rev. Microbiol.* **7,** 568–577 (2009).
24. Moran, J. J. *et al.* Methyl sulfides as intermediates in the anaerobic oxidation of methane. *Environ. Microbiol.* **10,** 162–173 (2008).
25. Stokke, R., Roalkvam, I., Lanzen, A., Haflidason, H. & Steen, I. H. Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ. Microbiol.* **14,** 1333–1346 (2012).
26. Nagarajan, H. *et al.* Characterization and modelling of interspecies electron transfer mechanisms and microbial community dynamics of a syntrophic association. *Nature Commun.* **4,** 2809 (2013).
27. Rotaru, A.-E. *et al.* A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to Methanosaeta for the reduction of carbon dioxide to methane. *Energ. Environ. Sci.* **7,** 408–415 (2014).
28. Malvankar, N. S. *et al.* Structural basis for metallic-like conductivity in microbial nanowires. *MBio* **6,** e00084–e00015 (2015).
29. Methé, B. A. *et al.* Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. *Science* **302,** 1967–1969 (2003).
30. Reitner, J. *et al.* Concretionary methane-seep carbonates and associated microbial communities in Black Sea sediments. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **227,** 18–30 (2005).

Supplementary Information is available in the online version of the paper.

Author Contributions G.W., V.K. and A.B designed the experiments; G.W., V.K., H.E.T. and D.R. performed experiments and data analyses; G.W., V.K. and A.B. wrote the manuscript with contributions from D.R. and H.E.T.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cultivation of TAOM consortia. Sediment-free, TAOM enrichment cultures were obtained after 1.5 years by semi-continuous incubation of hydrothermal vent sediments from Guaymas Basin with sulfate reducer medium[31] and 0.225 MPa CH_4 (+0.025 MPa CO_2) as the sole energy source at 60 °C, as described in ref. 6. Culture medium was replaced and samples were diluted 1:2 when sulfide concentrations exceeded 12 mM. For the different experiments, subsamples of the main culture (biological replicates) were incubated in parallel.

DNA extraction, sequencing and phylogenetic classification of TAOM partners. Genomic DNA was extracted as described previously[32] from an active TAOM culture. The protocol encompassed three cycles of freezing and thawing, chemical lysis in a high-salt extraction buffer (1.5 M NaCl) by heating of the suspension in the presence of sodium dodecyl sulfate and hexadecyltrimethylammonium bromide, and treatment with proteinase K. To amplify bacterial 16S ribosomal DNA genes the primer pair GM3/GM4 (ref. 33) was used. For archaeal 16S rDNA genes the primers 20F (ref. 34) and Arc1492R (ref. 35) were selected. PCR reactions were performed according to ref. 6. The phylogenetic affiliation was inferred with the ARB software package[36] and release 115 of the ARB SILVA database[37]. Representative 16S rRNA gene sequences are deposited at NCBI with the accession numbers KT152859–KT152885.

Visualization of TAOM aggregates by fluorescence *in situ* hybridization. Cell aliquots were fixed in 2% formaldehyde for 2 h at room temperature, washed with $1 \times PBS$ (pH 7.4). Fixed cell suspensions were treated with mild sonication (Sonoplus HD70; Bandelin) and aliquots of 50–250 μl were filtered onto GTTP filter (0.2 μm pore size, 20 mm diameter). CARD-FISH was performed as described previously[38] with the following modifications: for cell wall permeabilization, filters were sequentially incubated in lysozyme solution (10 mg ml^{-1} lysozyme powder, 0.1 M Tris–HCl, 0.05 M EDTA, pH 8) for 15–30 min at 37 °C and proteinase K solution (0.45 mU ml^{-1}, 0.1 M Tris–HCl, 0.05 M EDTA, pH 8, 0.5 M NaCl) for 2 min at room temperature. Endogenous peroxidases were inactivated by incubating the filters in 0.15% H_2O_2 in methanol (30 min, room temperature). The oligonucleotide probes ANME-1-350 and HotSeep-1-590 were applied with formamide concentrations according to ref. 6. For dual CARD-FISH, peroxidases of the first hybridization were inactivated by 0.3% H_2O_2 in methanol (30 min, room temperature). Catalysed reporter deposition was combined with the fluorochromes Alexa Fluor 488 and Alexa Fluor 594. Filters were stained with DAPI (4,6-diamidino-2-phenylindole). Micrographs were obtained by confocal laser scanning microscopy (LSM 780; Zeiss).

Test of potential AOM intermediates/alternative HotSeep-1 substrates. All experiments were performed with artificial seawater medium containing 30 mM of carbonate buffer at TAOM cultivation temperature (60 °C), except when specified otherwise. To ensure equilibration of gas phases, samples were agitated on shaking tables. Highly pure gases and chemicals were used as additions to the incubations. Standard TAOM conditions are defined here as 0.2 MPa methane and 28 mM sulfate. To test the TAOM enrichment for substrate-specific sulfide production, triplicate culture aliquots (10 ml in 20 ml Hungate tubes) were supplemented with different substrates (Extended Data Table 2) at concentrations of 20 mM, except methyl sulfide and carbon monoxide (both 0.05 MPa), and hydrogen (0.16 MPa) with and without methane (0.2 MPa). Zero-valent sulfur was prepared according to ref. 39 and was supplied as dissolved species. For this compound we additionally tested sulfide development via disproportion in a concentration gradient from 1–12 mM final S^0 concentration (Extended Data Fig. 6a). As positive reference, methane was provided at 0.2 MPa (at 60 °C roughly equivalent to 1.6 mM in solution). Sulfide production in the experiments was repeatedly measured every 3 to 4 days using the copper sulfide assay[40] and absorption spectrometry at 480 nm. TAOM rates with methane as the sole energy source (0.2 MPa) reached approximately 0.100 ± 0.030 μM per day, compared to a negative control (nitrogen; <0.001 μM per day). Rates determined for other substrates were compared to those under TAOM conditions.

Influence of hydrogen addition on methane oxidation in TAOM cultures. To determine the effect of hydrogen addition on methane oxidation rates, TAOM culture aliquots were supplemented with methane and hydrogen (0.15 MPa and 0.05 MPa, respectively), or only methane as control (0.15 MPa). Cultures were incubated headspace-free at 50 °C for this experiment, because hydrogen was too rapidly consumed at 60 °C for time-course experiments. To determine concentrations of methane and hydrogen, 1 ml of medium was sampled with gas-tight glass syringes, and the sampled medium was concurrently replaced with substrate-free medium to avoid the formation of a headspace. The sampled medium was

injected through the septum of 10 ml Exetainer filled with 1 ml NaOH and concentrations of CH_4 and H_2 were measured as described below.

Presence and production of hydrogen in active TAOM cultures. To determine hydrogen concentrations at TAOM conditions, 20 ml of culture was transferred into 156-ml bottles at 60 °C and gas phases were repeatedly sampled using glass syringes (1 ml) combined with direct measurements on the gas chromatograph. Cultures incubated for 3 or more days reached stable hydrogen concentrations. A comparison to molybdate addition is provided in Extended Data Fig. 6b, c. To quantify molecular hydrogen production in TAOM, 20 ml of culture was supplied with sodium molybdate (10 mM final concentration). This molybdate concentration assured complete inhibition of hydrogen-dependent sulfate reduction as shown in replicate incubations of TAOM culture (1 to 25 mM molybdate) with hydrogen (0.1 MPa) as the sole electron donor (Extended Data Fig. 6d). Samples were stored at 60 °C on a shaking table and repeatedly sampled by glass syringes. Concentrations of methane and hydrogen were measured via gas chromatography coupled to flame ionization detection (Focus GC, Thermo) and via reducing compound photometry (RCD; Peak Performer 1 RCP; Peak Laboratories).

Determination of carbon fixation by TAOM consortia. Replicate culture aliquots ($n = 5$) were incubated in 5-ml Hungate tubes supplemented with methane, sulfate and ^{14}C-labelled inorganic carbon (380 kBq). AOM-independent carbon fixation was determined under N_2 atmosphere ($n = 5$). To determine methane oxidation rates, replicate vials were incubated with ^{14}C-methane (14 kBq). Incubations were performed at 60 °C for 24 h. Samples were blotted onto 0.2-μm mixed cellulose esters membrane filters (Millipore, Merck). Filters were dried and potential residual inorganic carbon was removed by exposing the filters to an HCl atmosphere for 24 h. Radioactivity in liquid aliquots (0.1 ml) and filters was determined by liquid scintillation counting (scintillation mixture; Filtercount; Perkin Elmer; scintillation counter 2900TR LSA; Packard).

Cultivation of HotSeep-1 on molecular hydrogen. To isolate the hydrogenotrophic sulfate reducers in the TAOM enrichment, aliquots were transferred to Hungate tubes (20 ml) and diluted 1:10 to 1:10^9 with marine sulfate reducer medium. All vials were amended with 0.2 MPa H_2:CO_2 (80:20) gas phase, and additionally stored in N_2 atmosphere to prevent oxygen flux into the culture vials. Vials were stored at the TAOM temperature optimum (60 °C) and measured for sulfide production using the copper sulfate assay[40]. To identify cultivated microorganisms, the 16S rRNA gene of active hydrogenotrophic cultures was directly amplified from freeze-thawed pellets of culture aliquots (primer pair GM3/GM4) and sequenced as described above. The phylogenetic affiliation was inferred with the ARB software package[36] and Release 115 of the ARB SILVA database[37]. Representative sequences are deposited at NCBI with the accession numbers KT152886 and KT152887.

Physiology experiments with HotSeep-1. *Electron acceptor tests.* Culture aliquots (1 ml tenfold-diluted in artificial anoxic seawater medium) were supplied with different potential electron acceptors (colloidal sulfur, sulfite or thiosulfate) with and without the addition of hydrogen. Potential growth on alternative carbon sources (that is, acetate, butyrate, peptone and methyl sulfide) was tested.

Growth rates. Growth rates were independently determined from the development of sulfide concentrations and cell counts (from DAPI-stained cells for total cell numbers and from fluorescence *in situ* hybridized cells for specific cell numbers) from replicate cultures (grown from 10% inoculum).

Growth efficiencies. Efficiencies were determined in a ^{14}C-DIC radiotracer assay. Replicate cultures were spiked with ^{14}C-DIC ($\sim$ 5.4 MBq) and incubated with H_2:CO_2 or, as control, with N_2:CO_2 headspace. Sulfate-dependent hydrogen consumption was determined by the increase of sulfide (colourimetrically[40]) and by the decrease of sulfate (via ion chromatography) in the medium. Fixed carbon was measured from culture aliquots (5 ml volume) blotted on filters as described above. Concentrations of radioactivity on the filter and the medium were determined via scintillation counting. The total carbon fixation (mmol per ml culture) was calculated as ^{14}C uptake into particulate organic carbon multiplied by total DIC [^{14}C-POC (Bq per ml of culture)/^{14}C-total (Bq per ml of culture) $\times$ DIC (mmol per ml culture)], and normalized to reducing equivalent transfer, values are compared with the consumption of sulfide.

Metagenome sequencing and draft genome assembly of ANME-1 and HotSeep-1. Genomic DNA was extracted from TAOM and HotSeep-1 enrichment cultures (as described above) and prepared for Illumina sequencing using the Nextera mate pair sample preparation kit (Illumina), following the Gel-Plus protocol of the manufacturer's user guide. DNA fragments with a length of approximately 5–8 kb were extracted from a preparative gel before circularization. Additionally a paired-end read library with insert size of 500 bp was constructed for the TAOM enrichment using the TruSeq library preparation kit. Libraries were sequenced on a MiSeq instrument (MiSeq, Illumina) in a 2×250 bases paired-end run. Quality-controlled mate pair reads were assembled using the SPAdes genome

assembler v.3.5.0 (ref. 41) with default values of k and the -hqmp option. Assembled contigs from the TAOM metagenome were binned based on tetranucleotide frequency using the Metawatt software[42]. ANME-1- and HotSeep-1-specific bins were extracted for targeted reassembly using the SPAdes genome assembler v.3.5.0[41] with mapped mate pair and paired end read data and default values of k and subsequently were used as draft genomes. A HotSeep-1 draft genome was also obtained from the assembled contigs of the highly enriched HotSeep-1 culture metagenome (hydrogenotrophic HotSeep-1).

Draft genome analysis. Draft genomes were annotated with Prokka[43], and the draft genome of HotSeep-1 (obtained from the hydrogenotrophic HotSeep-1 enrichment) was additionally annotated with an in-house pipeline and analysed using GenDB[44] and JCoast[45]. The annotation of reported genes was manually curated. An expectation (E)-value cut-off of 1×10^{-5} was considered for all predictions of putative protein functions. Identity of the enriched hydrogenotrophic HotSeep-1 and the TAOM partner HotSeep-1 was evaluated by pairwise blast search of the nucleotide sequence of the 16S and 23S rRNA genes, functional and housekeeping genes and the intergenic spacer region (Extended Data Table 2) derived from the draft genome of the TAOM partner HotSeep-1 (query) versus the hydrogenotrophic HotSeep-1 (subject). To verify that the organisms belong to the same species the average nucleotide identity (ANI) and the tetranucleotide frequency correlation of the two draft genomes were determined using JSpecies[46] (v.1.2.1). Analyses resulted in tetranucleotide frequency correlation of 0.999 and ANI of >99%. To check for absence of ANME-1 in the hydrogenotrophic HotSeep-1 culture metagenomic reads were mapped to the SILVA SSU 119 reference database (bbmap v.35 and pyhloflash v.1.5) for phylogenetic classification at minimum identities of 90%, 95% and 97% resulting in approximately 3,500, 2,100 and 1,500 classified 16S rRNA gene fragments, respectively, which were screened for hits to ANME related sequences.

To identify potential cytochrome c and type IV pili (T4P) genes in the draft genomes of ANME-1 and HotSeep-1, protein domains were predicted using hmmscan (HMMER 3.0[47]) with the PfamA[48] and TIGRFAM[49] reference databases. Potential cytochromes were identified by the CXXCH motive and cytochrome c-specific protein domain models. Potential T4P genes were identified using protein models related to T4P. ANME-1 cytochrome and HotSeep-1 cytochrome and pili genes were compared for their best matching hits in the *G. sulfurreducens* (strain PCA) and *G. metallireducens* genome and the NCBI non-redundant protein database using blastp. Cytochrome annotation based on detected protein domains in PfamA, pili annotation based on detected protein domains and amino acid sequence. Subcellular localization was predicted with PSORTb[50] (v.3.0.2). For cytochromes the number of potential haem-binding sites was derived from the abundance of the CXXCH motif. For sequence comparison to the NCBI non-redundant protein database and *Geobacter* spp. and for details on protein domains and subcellular localization prediction see Supplementary Table 2a, b, (ANME-1 cytochromes), Supplementary Table 3a, b, (HotSeep-1 cytochromes) and Supplementary Table 4a, b (HotSeep-1 Type IV pili biogenesis). Representative sequences are deposited in GenBank under the accession numbers KT759143–KT759147, KT795302–KT795321, KT795322 and KT795323.

The ANME-1 draft genome was searched for genes encoding catalytic subunits of hydrogenases using blastp search against known genes of catalytic subunits of [NiFe] and [FeFe] hydrogenases (*mvhA, echA, frhA, vhuA, vhtA, ehaO, hymC*). Annotation of genes with hits was evaluated by blastp search against the NCBI non-redundant protein database for best matching reference sequences related to hydrogenases, but none were found.

Transcriptome analysis of TAOM and HotSeep-1. To collect cells for transcriptome analyses a 3.5-day experiment with replicates of 20 ml culture in 60-ml vials was carried out (Fig. 1). From triplicate TAOM cultures incubated with methane as control, with hydrogen, with methane/hydrogen mixture, or nitrogen as negative control, ~80% of the enrichment medium was removed and RNA was preserved using pre-heated RNA later (Life Technologies, ThermoFisher Scientific). Total RNA was extracted using the Quick-RNA MiniPrep kit (Zymo Research), treated with DNase I (Roche) and purified using the RNeasy MinElute Cleanup kit (Qiagen) following the manufacturer's recommendations. Removal of rRNA was omitted and total RNA was prepared for sequencing using the TruSeq stranded mRNA library prep kit (Illumina) following the manufacturer's guidelines. The cDNA library was sequenced on a MiSeq instrument (MiSeq, Illumina) generating between 2 to 3 million 150-bp single-end reads per library. Quality-controlled reads were mapped to the draft genomes of HotSeep-1 and ANME-1 using bbmap (v.35) with a minimum mapping identity of 98%. To quantify gene expression unambiguously mapped reads per gene were counted using bedtools multicov (v.2.24.0). To compare relative expression patterns within each organism across treatments, read counts per feature were converted to transcripts per million (TPM), which is the abundance of a specific gene (i) relative to the abundance

and length of all other transcribed genes (j) observed in one million sequenced reads calculated according to ref. 51:

$$TPM_i = X_i/l_i \times \left(1 \Big/ \sum_j X_j/l_j\right) \times 10^6$$

where X = counts and l = length (bp) per gene. Relative changes in expression of selected genes were calculated by comparing TPM normalized expression data of the H_2 and $H_2 + CH_4$ treatment to those under TAOM (control) conditions. Differential expression (P value, fold change and effect size) between control (TAOM condition) and treatment (H_2 or $H_2 + CH_4$) was computed with the aldex2 R package[52] for ANOVA-like differential expression analysis. Raw read numbers, read mapping data and statistical analysis are provided in Supplementary Table 1 (total expression) and Supplementary Table 5 (specific gene expression).

For HotSeep-1 transcriptomes total RNA was extracted from triplicate cultures (50 ml) grown on hydrogen/CO_2 following the same procedure as described for TAOM enrichments (see above). Removal of rRNA was omitted and total RNA was prepared for sequencing using the TruSeq stranded mRNA library prep kit (Illumina) following the manufacturer's guidelines. The cDNA library was sequenced on a MiSeq instrument (MiSeq, Illumina) generating between 6.4 to 6.9 million 75-bp paired-end reads per library. Quality-controlled reads were mapped to the draft genome of HotSeep-1 using bbmap (v.35) with a minimum mapping identity of 98%. To quantify gene expression unambiguously mapped reads per gene were counted using featureCount[53] (part of Subread, v.1.4.6.) with the -p option to count fragments instead of reads. Fragment counts per gene were converted to transcripts per million (TPM) as described above for TAOM transcriptome analyses.

Cultivation of *Geobacter* consortia. Active cultures of *G. sulfurreducens* (strain PCA; DSM 12127) and *G. metallireducens* (strain GS-15; DSM 7210) were mixed in fresh medium (DSM Medium 826) supplied with Na_2-fumarate (50 mM) and ethanol (20 mM) according to ref. 10 and cultivated anaerobically at 33 °C. After subsequent transfers (1% inoculum) a well-growing culture consisting of reddish microbial aggregates developed, which was used for thin sectioning and electron microscopy.

Transmission electron microscopy. The cell material was harvested at 2,000 r.p.m. using a Stat Spin Microprep 2 table-top centrifuge. After centrifugation the pellet was fixed by immersion using 2% glutaraldehyde in 0.1 M cacodylate buffer at pH 7.4. Fixation was performed for 60 min at room temperature. The fixed pellet was immobilized with 2% agarose in cacodylate buffer at pH 7.4. The pellet was cubed and the pieces carefully washed with buffer and further fixed in 1% osmium tetroxide. After pre-embedding staining with 1% uranyl acetate, samples were dehydrated and embedded in Agar 100 (Epon 812 equivalent). As an independent complementary method (shown in Extended Data Fig. 5a), samples were placed in aluminium platelets of 150 μm depth containing 1-hexadecene (ref. 54). The platelets were frozen using a Leica Em HPM100 high pressure freezer (Leica Mikrosysteme Vertrieb GmbH). The frozen samples were transferred to an Automatic Freeze Substitution Unit Leica EM AFS2 and substituted at −90 °C in a solution containing anhydrous acetone, 0.1% tannic acid for 24 h and in anhydrous acetone, 2% OsO$_4$, 0.5% anhydrous glutaraldehyde (EMS Electron Microscopical Science) for an additional 8 h. After a further incubation over 20 h at −20 °C samples were warmed up to + 4 °C and washed with anhydrous acetone subsequently. The samples were embedded at room temperature in Agar 100 at 60 °C over 24 h. Thin sections (30–60 nm) were counterstained with uranyl acetate and lead citrate and examined using a Philips CM 120 transmission electron microscope (Philips Inc.). In total, we recorded more than 200 views on TAOM consortia, 64 views on HotSeep-1 and 90 views of *Geobacter* consortia.

Thermodynamic calculations. Free energy yields (ΔG_{rxn}) were calculated according to the equation:

$$\Delta G_{rxn} = \Delta G^{'}_{(T)} + RT \times \ln \frac{(P_i)^n}{(R_i)^n}$$

including the gas constant R, the temperature T (K) and the measured activities/partial pressures of the respective products P_i and reactants R_i in their respective stoichiometric appearance (n) in the reaction. Values consider activities and fugacity of respective compounds. The temperature-corrected standard free energy $\Delta G^{'}_{(T)}$ were determined according to ref. 55.

31. Widdel, F. & Bak, F. in *The Prokaryotes* Vol. 4 (eds Trüper, H. G., Balows, A., Dworkin, M., Harder, W. & Schleifer, K. H.) 3352–3378 (Springer, 1992).

32. Zhou, J., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).

33. Muyzer, G., Teske, A., Wirsen, C. O. & Jannasch, H. W. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent

samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch. Microbiol.* **164,** 165–172 (1995).

34. Massana, R., Murray, A., Preston, C. & DeLong, E. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl. Environ. Microbiol.* **63,** 50–56 (1997).
35. Teske, A. *et al.* Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl. Environ. Microbiol.* **68,** 1994–2007 (2002).
36. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
37. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
38. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68,** 3094–3101 (2002).
39. Steudel, R., Göbel, T. & Holdt, G. The molecular composition of hydrophilic sulfur sols prepared by decomposition of thiosulfate. *Z. Naturforsch. B Chem. Sci.* **43,** 203–218 (1988).
40. Cord-Ruwisch, R. A quick method for the determination of dissolved and precipitated sulfides in cultures of sulfate-reducing bacteria. *Microbiol. Meth.* **4,** 33–36 (1985).
41. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
42. Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3,** 410 (2012).
43. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
44. Meyer, F. *et al.* GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **31,** 2187–2195 (2003).
45. Richter, M. *et al.* JCoast — a biologist-centric software tool for data mining and comparison of prokaryotic (meta) genomes. *BMC Bioinformatics* **9,** 177 (2008).
46. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl Acad. Sci. USA* **106,** 19126–19131 (2009).
47. Eddy, S. *HMMER User's Guide. Biological sequence analysis using profile hidden Markov models* (Howard Hughes Medical Institute, 2003).
48. Finn, R. D., Miller, B. L., Clements, J. & Bateman, A. iPfam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Res.* **42,** D364–D373 (2014).
49. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31,** 371–373 (2003).
50. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26,** 1608–1615 (2010).
51. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26,** 493–500 (2010).
52. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2,** 1–13 (2014).
53. Liao, Y., Smyth, G. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
54. Studer, D., Michel, M. & Müller, M. High pressure freezing comes of age. *Scanning Microsc.*, Suppl. 3, 253–268 (1989).
55. Conrad, R. & Wetter, B. Influence of temperature on energetics of hydrogen metabolism in homoacetogenic, methanogenic, and other anaerobic bacteria. *Arch. Microbiol.* **155,** 94–98 (1990).

a

H_2S

$4H_2$

SO_4^{2-}

CO_2

$8e^-$ $8H^+$

CH_4

b

$7\ SO_4^{2-}$

$7\ H_2S$

$7\ CO_2$

S_O

$4HS_2^-$

$7\ CH_4$

c

H_2S

$8e^-$

SO_4^{2-}

CO_2

$8e^-$

Nanowires
(or direct contact)

CH_4
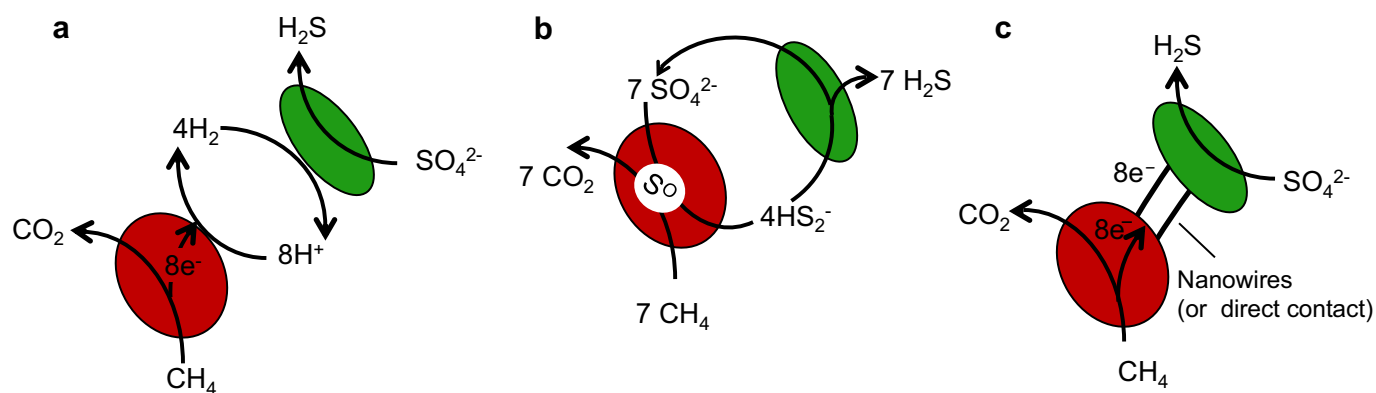
Extended Data Figure 1 | Models of possible species interaction mechanisms in TAOM tested in this study. **a**, Transfer of molecular intermediates such as hydrogen. **b**, Incomplete reduction of sulfate in ANME and zero-valent sulfur transfer to the partner bacteria. **c**, Direct interspecies electron transfer via conductive nanowires.
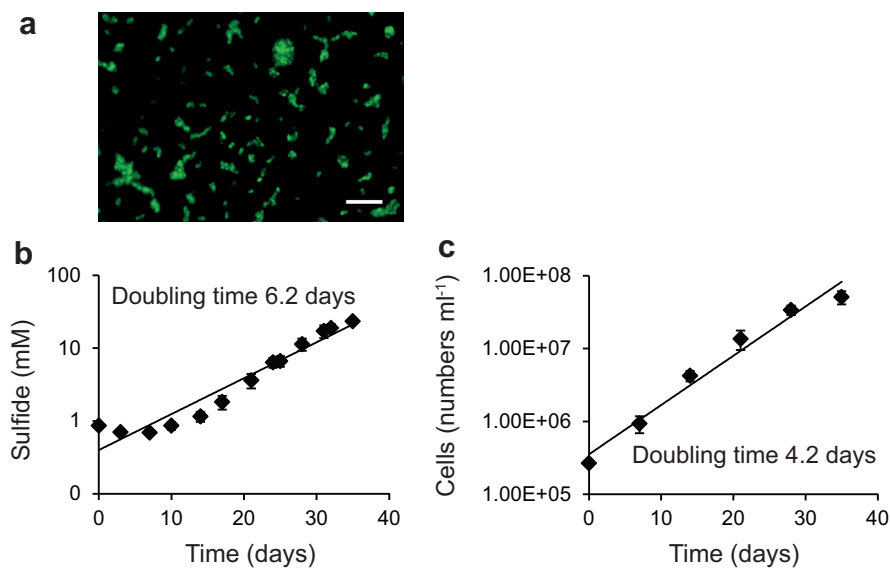
Extended Data Figure 2 | Visualization of and growth experiments with HotSeep-1. **a**, Representative fluorescence micrograph of HotSeep-1 culture (probe HotSeep-1-590; 22 similar images obtained). Cells are solitary or form small aggregates. Scale bar, 10 μm. **b**, **c**, Semi-logarithmic illustration of the development of sulfide (**b**) or numbers of cells and resulting doubling times (**c**) (doubling time = ln(2)/exponential factor of the regression curve) during incubation of the HotSeep-1 culture with hydrogen as the sole energy source and sulfate. Biological replicates $n = 3$; data is presented as mean ± s.d., lines of best fit defined by least squares method.
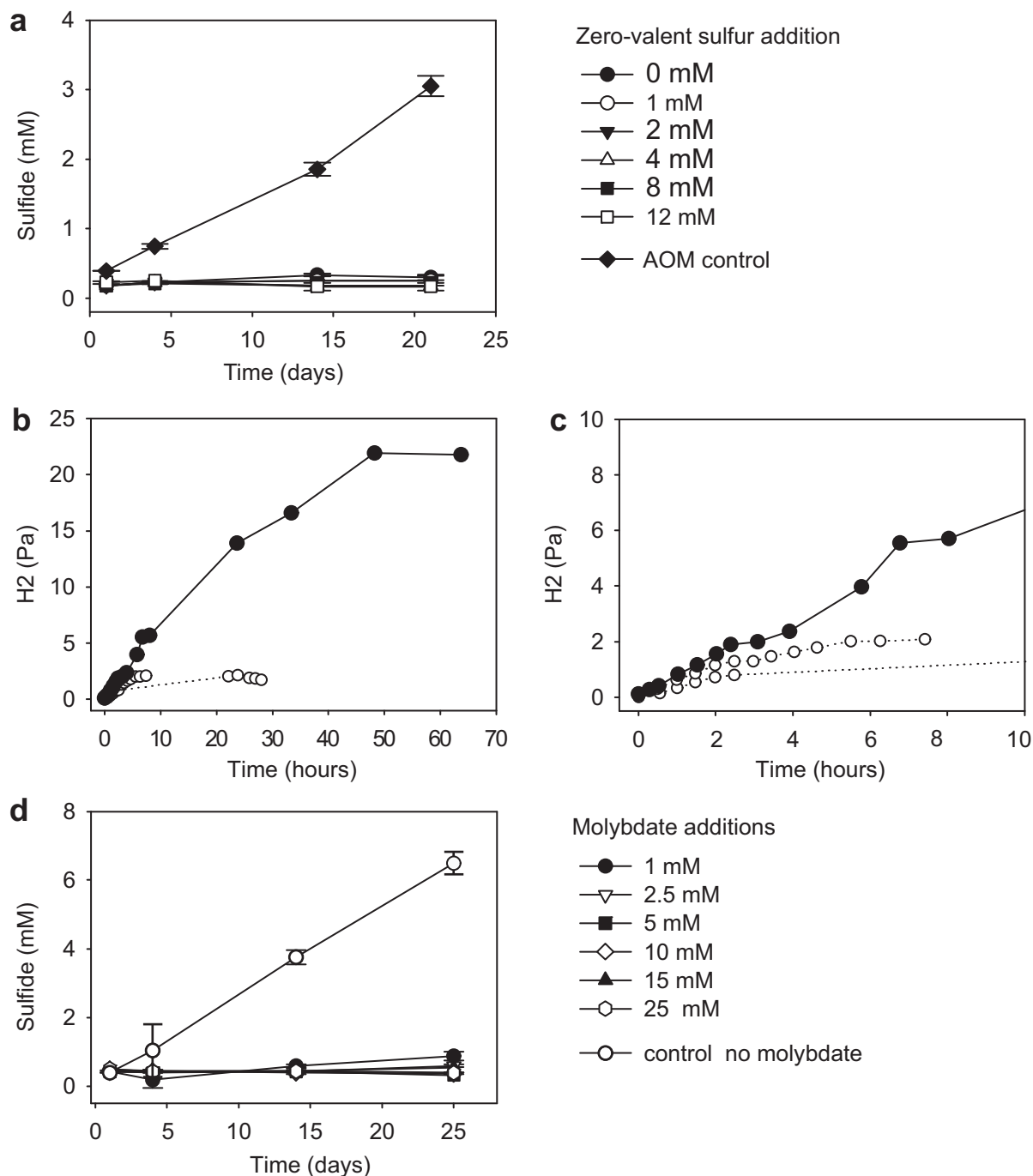
Extended Data Figure 3 | Effect of zero-valent elemental sulfur and molybdate additions on TAOM. **a**, Sulfide production in response to zero-valent (colloidal) sulfur addition versus TAOM conditions; zero-valent sulfur did not cause sulfide formation. **b**, **c**, Monitoring of hydrogen partial pressures at TAOM conditions (open circles) versus extra addition of 10 mM molybdate (filled circles) for either the full times series (**b**) or the first 10 h

(**c**). Molybdate addition caused tenfold higher hydrogen concentrations than the TAOM condition. **d**, Inhibition of methane-dependent sulfide production at different molybdate concentrations. Biological replicates $n = 3$; symbols represent mean values; error bars are s.d.; **b** and **c** show a single time series with the same culture.
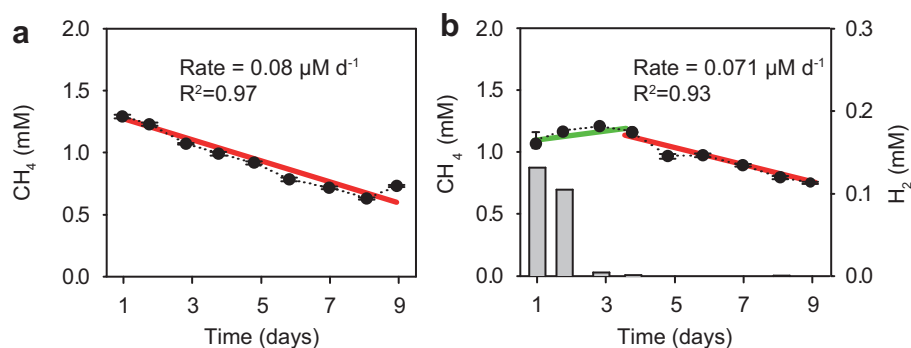
Extended Data Figure 4 | Effect of hydrogen on microbial methane oxidation. **a**, Methane (0.15 MPa) supplied as the sole electron source was steadily consumed over time by TAOM. **b**, When both methane (0.15 MPa) and hydrogen (0.05 MPa) were added, hydrogen was rapidly consumed (grey bars), whereas methane consumption was reversely inhibited (green line) until hydrogen was fully consumed. Afterwards methane consumption occurred at the same rate as in the control with only methane (**a**). Methane, technical replicates $n = 3$; symbols represent mean values; error bars are s.d.; hydrogen, single measurements. Experiment was replicated once in the laboratory.

Relative expression of genes of HotSeep-1
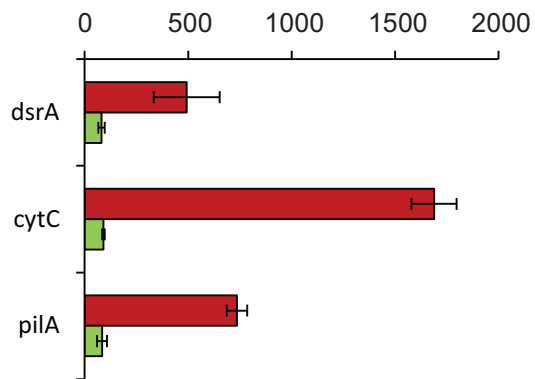
Extended Data Figure 5 | Relative expression of marker genes of HotSeep-1 in consortial growth on methane (TAOM) versus enrichment on hydrogen. Genes encoding proteins apparently involved in direct interspecies electron transfer (CytC and PilA) are strongly overexpressed during TAOM (red) compared to hydrogenotrophic growth (green). Gene expression given in terms of TPM (transcripts per million). Biological replicates $n = 3$; error bars are s.d.
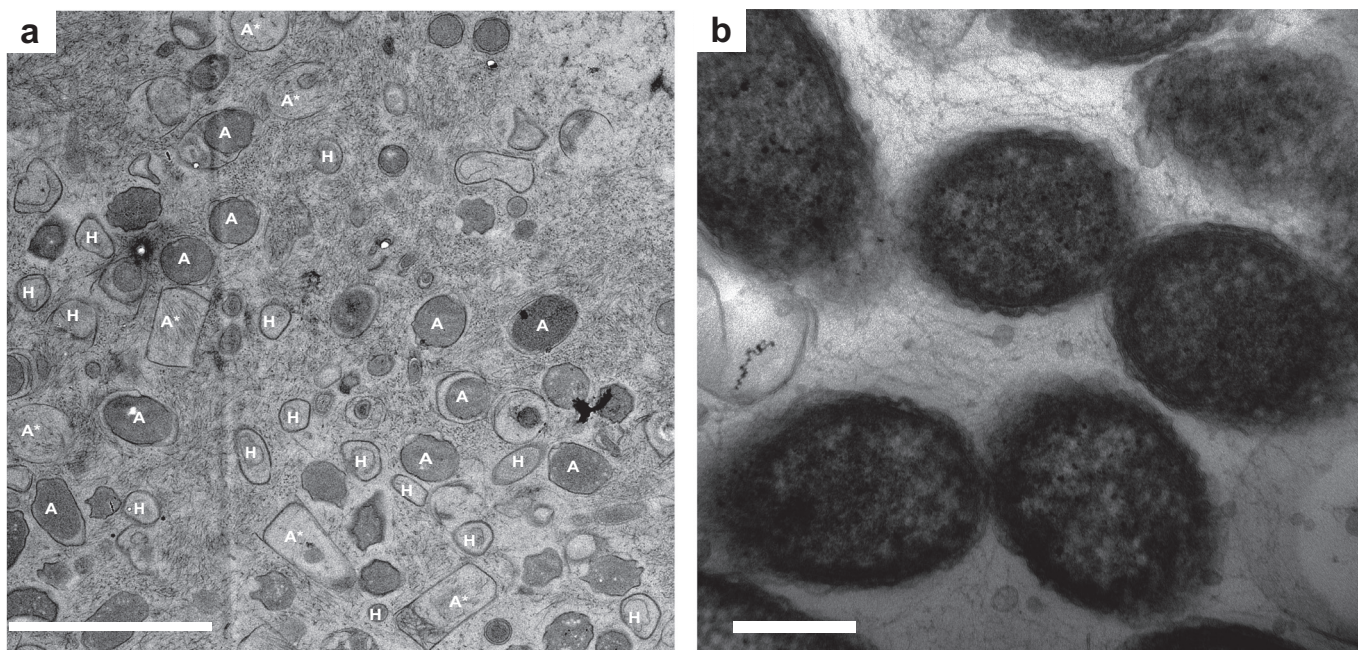
Extended Data Figure 6 | Thin-sections of TAOM and dual species
***Geobacter* spp. aggregates. a**, TAOM culture. High-pressure frozen ANME-1
cells (A) have a cylindrical shape and a size of $1.5 \times 0.8\,\mu m$, appearing
circular in cross-section, and rectangular when cut along the axis. Their cell
content shows a high contrast. A* indicates cell envelopes. HotSeep-1 cells (H)
are smaller (approximately $1 \times 0.5\,\mu m$), of rod-like shape, and have lower
contrast. The matrix between the cells is largely filled with filaments.
Representative of 24 images. Scale bar, $3\,\mu m$. **b**, Thin section of *Geobacter*
consortium with intercellular nanowires using the same embedding techniques
as for TAOM consortia, representative of 20 images. Scale bar, 300 nm.

Extended Data Table 1 | Phylogenetic affiliation of cloned 16S rRNA gene sequences obtained from TAOM enrichments in 2010 (compiled from ref. 6) and after 1.5 years of cultivation in 2012 (this study)

Phylogenetic Group	No of clones	
	2010, slurry	2012, sediment-free
Archaea		
Euryarchaeota		
ANME-1		
ANME-1-Guaymas cluster	**46 (82%)**	**148 (89%)**
Other related ANME-1	-	7
Thermoplasmatales (19c-33 related)	6	7
Thermococcales	-	1
Others	4	3
Sum	**56**	**166**
Bacteria		
Proteobacteria		
Betaproteobacteria	3	1
Gammaproteobacteria	11	-
Deltaproteobacteria		
HotSeep-1-Cluster	**124 (60%)**	**89 (59%)**
DSS group	1	1
Others	16	4
Acidobacteria	1	-
Actinobacteria	3	-
Candidate Division OD1	6	1
Candidate Division OP3	5	40
Candidate Division OP8	21	6
Chloroflexi	8	-
Others	8	9
Sum	**207**	**151**

Extended Data Table 2 | Pairwise comparison of nucleotide sequences from the HotSeep-1 draft genomes derived from the TAOM culture versus the HotSeep-1 culture with hydrogen

Feature	Identity (bp)	Identity (%)	Gaps	Query coverage	E value
16S rRNA	1554/1555	99	0/1555	100	0
23S rRNA	3025/3029	99	0/2029	100	0
ITS (23S-16S)	270/271	99	0/271	100	3.00E-145
dsrA	1406/1438	98	0/1438	100	0
aprA	1905/1905	100	0/1905	100	0
Hydogenase small subunit	1437/1437	100	0/1437	100	0
Hydrogenase large subunit	916/918	100	0/918	99	0
dnaK	1884/1893	99	0/1893	100	0

ITS, internal transcribed spacer.

Extended Data Table 3 | Effect of potential intermediates in AOM on sulfide production of TAOM culture (*n* = 3 replicates, 20 days incubation)

Substrate	Sulfide production	
	plus substrate	plus methane
Control (no donor)	−	
Methane	+	
Colloidal sulfur	−	+
Hydrogen	+++	+++
Carbon monoxide	−	−
Methyl sulfide	−	−
Methanol	−	+
Acetate	−	+
Formate	−	+

−, sulfide production at level of negative control; +, sulfide production similar to TAOM under standard conditions; +++, sulfide production tripled compared to TAOM standard conditions.

Extended Data Table 4 | Genes encoding cytochrome *c* proteins identified in thermophilic ANME-1 and HotSeep-1 draft genomes, and for type IV pili biogenesis identified in the HotSeep-1 draft genome with expression >20 transcripts per million.

Cytochrome c type based on PfamA domain prediction	Predicted cellular localization (PSORTb)	Predicted heme groups	Expression in TAOM (TPM)	Expression change in H₂ treatment
ANME-1				
Cytochrom_c3_2	**Unknown (CM,CW,E)**	**8**	1,063	↓
Cytochrome_C7	Unknown (CM,CW,E)	4	603	↓
Cytochrome_C7	Extracellular	4	506	↓
Cytochrom_NNT	Cytoplasmic	5	73	↓
HotSeep-1				
Paired_CXXCH_1	**Extracellular**	**6**	2,485	↓
Cytochrom_CIII	Periplasmic	4	1,011	-
Paired_CXXCH_1	Unknown (CM,OM,P,E)	7	974	-
Cytochrome_C554	Unknown (CM,OM,P,E)	5	881	-
Cytochrom_CIII	Periplasmic	4	179	-
Cytochrome_C7	Cytoplasmic Membrane	5	95	-
Paired_CXXCH_1	Cytoplasmic	10	95	-
Cytochrom_c3_2	Unknown (CM,P,E)	12	86	-
Cytochrom_c3_2	Periplasmic	12	74	-
Cytochrome_C554	Unknown (CM,P,E)	4	24	↓

Predicted pili protein	Predicted cellular localization (PSORTb)	% identity to *G. sulfurreducens*	Expression in TAOM (TPM)	Expression change in H₂ treatment
HotSeep-1				
assembly protein (*pil*A)	**Extracellular**	**74**	**1084**	↓
retraction ATPase (*pil*T)	Cytoplasmic	40	51	↓
assembly protein (*pil*Y)	Extracellular	26	46	↓
assembly ATPase (*pil*B)	Cytoplasmic	47	26	-
secretion (*pil*Q)	Unknown (OM, C)	32	26	-
assembly protein (*pil*A)	Cytoplasmic Membrane	41	23	↓
retraction ATPase (*pil*T)	Cytoplasmic	55	21	-
assembly protein (*pil*M)	Cytoplasmic	35	21	-
assembly protein (*pil*O)	Cytoplasmic Membrane	35	21	↓

Genes in bold are presented in Fig. 3a. CM, cytoplasmic membrane; CW, cell wall; E, extracellular; OM, outer membrane; P, periplasm; TPM, transcripts per million; ↑, upregulated by a factor of 2; ↓ downregulated by a factor of 2; –,change smaller than by a factor of 2.

LETTER

Dynamic m^{6}A mRNA methylation directs translational control of heat shock response

Jun Zhou[1], Ji Wan[1], Xiangwei Gao[1], Xingqian Zhang[1], Samie R. Jaffrey[2] & Shu-Bing Qian[1]

The most abundant mRNA post-transcriptional modification is N^6-methyladenosine (m^{6}A), which has broad roles in RNA biology[1–5]. In mammalian cells, the asymmetric distribution of m^{6}A along mRNAs results in relatively less methylation in the 5′ untranslated region (5′UTR) compared to other regions[6,7]. However, whether and how 5′UTR methylation is regulated is poorly understood. Despite the crucial role of the 5′UTR in translation initiation, very little is known about whether m^{6}A modification influences mRNA translation. Here we show that in response to heat shock stress, certain adenosines within the 5′UTR of newly transcribed mRNAs are preferentially methylated. We find that the dynamic 5′UTR methylation is a result of stress-induced nuclear localization of YTHDF2, a well-characterized m^{6}A 'reader'. Upon heat shock stress, the nuclear YTHDF2 preserves 5′UTR methylation of stress-induced transcripts by limiting the m^{6}A 'eraser' FTO from demethylation. Remarkably, the increased 5′UTR methylation in the form of m^{6}A promotes cap-independent translation initiation, providing a mechanism for selective mRNA translation under heat shock stress. Using Hsp70 mRNA as an example, we demonstrate that a single m^{6}A modification site in the 5′UTR enables translation initiation independent of the 5′ end N^7-methylguanosine cap. The elucidation of the dynamic features of 5′UTR methylation and its critical role in cap-independent translation not only expands the breadth of physiological roles of m^{6}A, but also uncovers a previously unappreciated translational control mechanism in heat shock response.

Given the reversible nature of m^{6}A mRNA methylation[8,9], we sought to assess the potential impact of heat shock stress on m^{6}A modification of eukaryotic mRNAs. Using immunofluorescence staining, we first examined the subcellular localization of the entire m^{6}A machinery in a mouse embryonic fibroblast (MEF) cell line before and after heat shock stress. It is believed that m^{6}A modification occurs primarily at nuclear speckles, whereas its functionality takes place in the cytosol (Fig. 1a). Consistent with this notion, both the m^{6}A 'writers' (METTL3, METTL14, WTAP) and the eraser FTO were predominantly present in the nucleus, whereas the majority of the reader YTHDF2 resided in the cytosol (Fig. 1b and Extended Data Fig. 1). In response to heat shock stress, neither the writers nor the eraser changed their nuclear localization (Extended Data Fig. 1). Surprisingly, nearly all of the YTHDF2 molecules were relocated into the nucleus from the cytosol upon heat shock stress (Fig. 1b). The same phenomenon holds true in HeLa cells. Intriguingly, the protein level of YTHDF2 was also markedly increased after heat shock stress in a manner similar to Hsp70 induction (Fig. 1c). In contrast, neither the m^{6}A writers nor the eraser showed any differences in protein levels upon stress. Supporting the stress-induced transcriptional upregulation of YTHDF2, real-time PCR revealed a nearly fourfold increase of mRNA abundance after heat shock stress (Fig. 1d). The increased YTHDF2 abundance was not due to altered mRNA degradation since heat shock stress had negligible effects on mRNA stability (Extended Data Fig. 2a). Notably, YTHDF2 exhibited a relatively short half-life ($t_{1/2} < 1$ h) in cells, supporting the importance of stress-induced

transcriptional upregulation. Genes encoding other YTH domain family proteins like YTHDF1 and YTHDF3 also showed upregulation, although to a lesser extent (Extended Data Fig. 2b). Using a mouse fibroblast cell line lacking the heat shock transcription factor 1 (HSF1)[10], we confirmed that YTHDF2 is subject to regulation by HSF1 (Extended Data Fig. 2c). The unexpected stress-inducible feature of YTHDF2 suggests a potential role of m^{6}A modification in heat shock response.

Although YTHDF2 primarily serves as the reader of m^{6}A, recent proteomic data revealed that YTHDF2 has an extensive physical interaction with the components of m^{6}A writers[11]. Given their co-localization upon heat shock stress, we postulated that the nuclear presence of YTHDF2 could influence the m^{6}A modification and alter the landscape of mRNA methylomes. Using an optimized m^{6}A-seq procedure[6,12], we sequenced the entire methylated RNA species purified from MEF cells with or without heat shock stress. From a total of 15,454 putative methylation sites, we confirmed the m^{6}A consensus sequence motif as GGAC (where the underlined A is modified) (Extended Data Fig. 3 and Supplementary Table 1). Consistent with previous reports[6,7], the majority of m^{6}A sites are enriched in the
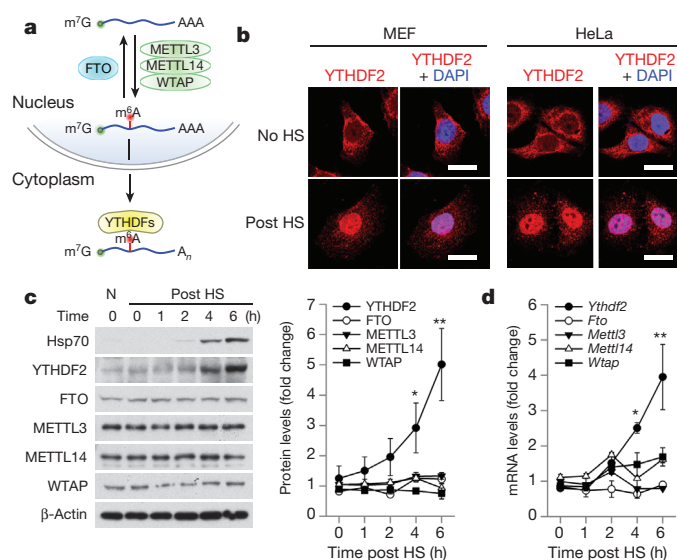
Figure 1 | YTHDF2 changes cellular localization and expression levels in response to heat shock stress. a, Schematic of m^{6}A modification machinery in mammalian cells. **b**, Subcellular localization of YTHDF2 in MEF and HeLa cells before or 2 h after heat shock (42 °C, 1 h). Bar, 10 μm. Images are representative of at least 50 cells. **c**, Immunoblotting of MEF cells after heat shock stress (42 °C, 1 h). N, no heat shock. The right panel shows the relative protein levels quantified by densitometry and normalized to β-actin. Representative of three biological replicates. **d**, Same samples in **c** were used for RNA extraction and real-time PCR. Relative levels of indicated transcripts are normalized to β-actin. Error bars, mean ± s.e.m.; *$P < 0.05$, **$P < 0.01$, unpaired two-tailed t-test; $n = 3$ biological replicates (**c** and **d**).

[1]Division of Nutritional Sciences, Cornell University, Ithaca, New York 14853, USA. [2]Department of Pharmacology, Weill Cornell Medical College, Cornell University, New York City, New York 10065, USA.

Figure 2 | Altered m⁶A profiles in MEF cells in response to heat shock stress.
a, Metagene profiles of m⁶A distribution across the transcriptome of cells before or 2 h after heat shock (42 °C, 1 h). Black arrow indicates the m⁶A peak in the 5′UTR region. **b**, Transcripts are stratified by different expression levels after heat shock stress, followed by metagene profiles of m⁶A distributio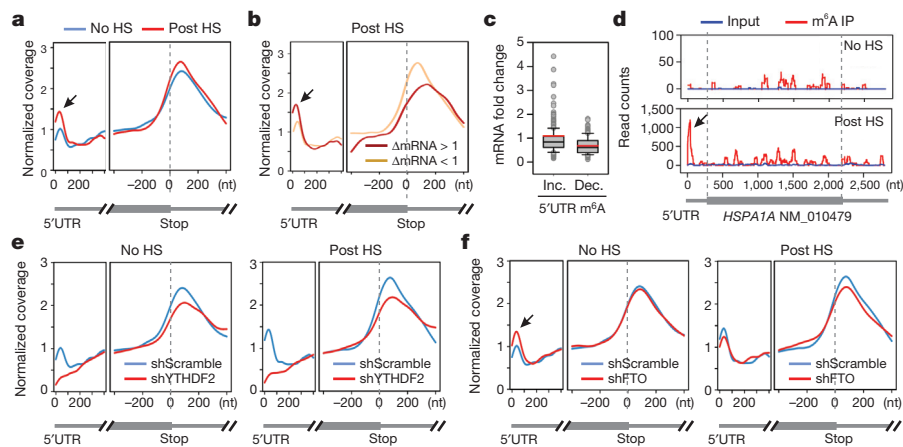n. **c**, A box plot depicting fold changes of mRNA levels after heat shock for transcripts showing increased or decreased m⁶A modification in the 5′UTR.

Box plot centre line (black), mean; whiskers, 5th and 95th percentiles; red line, median. **d**, An example of stress-induced transcript *HSPA1A* harbouring m⁶A peaks. IP, immunoprecipitation. **e**, Metagene profiles of m⁶A distribution across the transcriptome of cells with or without YTHDF2 knockdown, before or after heat shock stress. **f**, Metagene profiles of m⁶A distribution across the transcriptome of cells with or without FTO knockdown, before or after heat shock stress.

vicinity of the stop codon and in the 3′UTR (Fig. 2a). Unexpectedly, heat shock stress led to an elevated m⁶A peak in the 5′UTR, but not other regions. Reasoning that only a handful of genes undergo upregulation as a result of heat shock response[13,14], we compared the levels of m⁶A modification between stress-inducible and non-inducible transcripts defined by RNA-seq. It is clear that the upregulated transcripts showed greater m⁶A modification in the 5′UTR than the transcripts downregulated upon stress (Fig. 2b). We next stratified transcripts based on differential changes of m⁶A modification in the 5′UTR. While transcripts with elevated 5′UTR methylation are mostly upregulated in response to stress, transcripts exhibiting decreased 5′UTR methylation are largely downregulated (Fig. 2c, $P < 0.001$, Mann–Whitney Test). One particular example of stress-induced transcripts is the Hsp70 gene *HSPA1A*, which not only showed a 90-fold increase of mRNA levels after heat shock, but also displayed a prominent m⁶A peak in the 5′UTR (Fig. 2d). By contrast, the constitutively expressed Hsc70 gene *HSPA8* showed only minor increase in both the mRNA level and the m⁶A modification in response to heat shock stress (Extended Data Fig. 4). These results suggest that the increased 5′UTR methylation selectively occurs on the stress-inducible mRNAs.

To examine whether the elevated 5′UTR methylation upon heat shock stress is a result of nuclear localization of YTHDF2, we silenced YTHDF2 in MEF cells using lentiviruses expressing short hairpin RNAs. Remarkably, MEF cells lacking YTHDF2 demonstrated a substantial loss of m⁶A modification in the 5′UTR (Fig. 2e). Upon heat shock stress, these cells no longer showed the elevated 5′UTR methylation as seen in control cells. The abolished 5′UTR methylation in the absence of YTHDF2 was clearly exemplified in *HSPA1A* that exhibited only background m⁶A modification in the 5′UTR (Extended Data Fig. 5). This result indicates a novel function of YTHDF2 in heat shock response by promoting 5′UTR methylation on mRNAs transcribed during stress.

YTHDF2 is not a methyltransferase *per se*, and does not bind to mRNAs without prior m⁶A modification[3,7]. How does the nuclear presence of YTHDF2 promote selective methylation in the 5′UTR? One possibility is that YTHDF2 protects the pre-existing m⁶A from FTO-mediated demethylation. Upon heat shock stress, the nuclear localization of YTHDF2 probably limits the accessibility of FTO to newly minted m⁶A sites, thereby tilting the equilibrium towards methylation. Indeed, an *in vitro* m⁶A binding and demethylation assay confirmed direct competition between FTO and YTHDF2 (Extended Data Fig. 6). To investigate whether FTO preferentially removes m⁶A

modification from the 5′UTR, we knocked down FTO from MEF cells and examined the m⁶A distribution across the entire transcriptome. Notably, only the 5′UTR region showed an increase of m⁶A density in cells lacking FTO (Fig. 2f). Additionally, the 5′UTR methylation showed no further increase upon heat shock stress in the absence of FTO.

The 5′UTR is crucial in mediating translation initiation of eukaryotic mRNAs[15,16]. Under stress conditions, the cap-dependent translation is generally suppressed. However, subsets of transcripts are selectively translated via a poorly understood cap-independent mechanism[17–19]. To investigate whether differential methylation of 5′UTR influences the translational status of these mRNAs, we conducted ribosome profiling of MEF cells with or without heat shock stress. Among the genes undergoing stress-induced transcriptional upregulation, many not only showed elevated m⁶A modification in the 5′UTR, but also demonstrated increased ribosome occupancy in the coding region (Fig. 3a). Several prominent examples are genes encoding heat shock proteins, in particular Hsp70 (Supplementary Table 2). Therefore, the coordinated upregulation of transcription and 5′UTR methylation is coupled with robust translation in response to heat shock stress.

To validate the causal relationship between stress-induced 5′UTR methylation and selective translation, we examined Hsp70 synthesis in cells with differential m⁶A modification. Knocking down YTHDF2 leads to depleted 5′UTR methylation, as revealed by m⁶A-seq (Fig. 2e). Indeed, direct m⁶A blotting of *HSPA1A* purified from heat-shock-stressed MEFs confirmed the marked reduction of methylation in cells lacking YTHDF2 (Fig. 3b). Remarkably, the heat-shock-induced Hsp70 synthesis was substantially reduced in the absence of YTHDF2 (Fig. 3c). The comparable Hsp70 mRNA levels in cells with or without YTHDF2 knockdown indicate that the reduced Hsp70 synthesis is a result of translational deficiency (Extended Data Fig. 7). Further supporting this notion, the Hsp70 transcript, but not GAPDH, showed much less enrichment in the polysomes of MEF cells lacking YTHDF2 (Fig. 3d).

Reasoning that YTHDF2 competes with FTO in preserving 5′UTR m⁶A modification, we speculated that FTO knockdown would increase the 5′UTR methylation as well as the translation efficiency of Hsp70 mRNA. This was indeed the case. Direct m⁶A blotting of *HSPA1A* purified from stressed MEFs lacking FTO revealed a clear increase of methylation when compared to the scramble control (Extended Data Fig. 8). Importantly, FTO knockdown potentiated
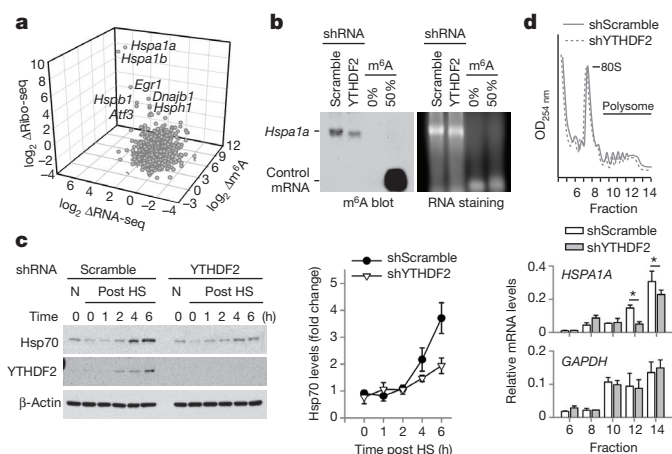
Figure 3 | m⁶A modification promotes selective translation under heat shock stress. **a**, A 3D plot depicting fold changes (log₂) of mRNA abundance, coding sequence ribosome occupancy (Ribo-seq), and 5′UTR m⁶A levels in MEF cells after heat shock stress. **b**, m⁶A blotting of *HSPA1A* purified from MEFs with or without YTHDF2 knockdown. Messenger RNAs synthesized by *in vitro* transcription in the absence or presence of m⁶A were used as control. Images are representative of two biological replicates. **c**, Immunoblotting of MEF cells with or without YTHDF2 knockdown after heat shock stress (42 °C, 1 h). N, no heat shock. The right panel shows the relative protein levels quantified by densitometry and normalized to β-actin. Blots are representative of three biological replicates. **d**, MEF cells with or without YTHDF2 knockdown were subject to heat shock stress followed by sucrose gradient sedimentation. Specific mRNA levels in polysome fractions were measured by quantitative PCR. The values are first normalized to the spike in control then to the total. Error bars, mean ± s.e.m.; *P < 0.05, unpaired two-tailed *t*-test; n = 3 biological replicates (**c** and **d**).

the synthesis of Hsp70 after heat shock stress. Collectively, these results established the functional connection between dynamic 5′UTR methylation and selective mRNA translation during stress.

It is commonly believed that the 5′UTR of Hsp70 mRNA recruits the translational machinery via an internal ribosome entry site (IRES)[20–23]. However, conflicting results exist and the exact cap-independent translation-promoting determinants remain elusive[23,24]. Given the fact that the normal 5′ end cap structure is a methylated purine (N⁷-methylguanosine, m⁷G), we hypothesize that the stress-induced m⁶A in the 5′UTR enables selective translation by acting as a functional cap substitute. To test this hypothesis, we performed a firefly luciferase (Fluc) reporter assay in MEF cells by transfecting mRNAs synthesized in the absence or presence of m⁶A (Fig. 4a). For the messenger without 5′UTR, random incorporation of m⁶A slightly reduced the Fluc activity after mRNA transfection. In the presence of 5′UTR from Hsp70, but not tubulin, the incorporation of m⁶A markedly increased the Fluc activity in transfected MEF cells. Notably, m⁶A incorporation does not affect the stability of the synthesized mRNAs in transfected cells (Extended Data Fig. 9a). We next replaced the 5′ end m⁷G cap with a non-functional cap analogue ApppG. As expected, the resultant mRNA did not support translation in the absence of 5′UTR or in the presence of tubulin 5′UTR (Fig. 4a and Extended Data Fig. 9b). Only when the Hsp70 5′UTR was present, was the translating-promoting feature clearly manifested after m⁶A incorporation, in particular under stress conditions (Fig. 4a). This effect is specific to m⁶A modification but not m⁶Am because ribose methylation in the form of 2′-O-MeA suppressed translation of the Fluc reporter bearing Hsp70 5′UTR (Fig. 4a). Therefore, methylation of Hsp70 5′UTR in the form of m⁶A promotes cap-independent translation.

To further demonstrate the 5′UTR specificity in m⁶A-facilitated cap-independent translation, we examined 5′UTRs from a constitutively expressed chaperone Hsc70 (*HSPA8*) and another stress-inducible chaperone Hsp105 (*HSPH1*) (Fig. 3a). Only the 5′UTR of
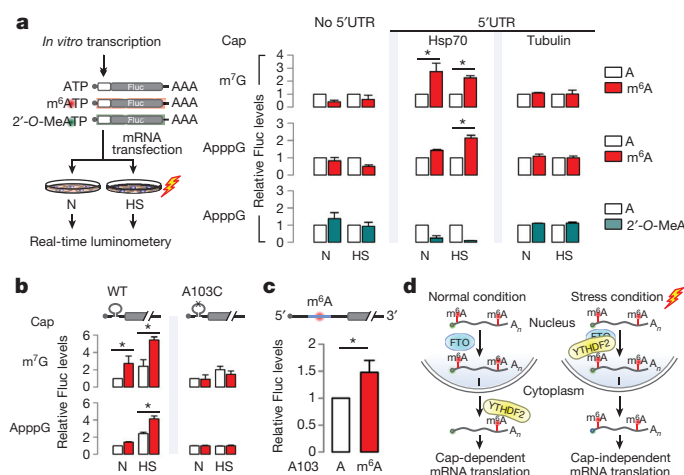

Figure 4 | Selective 5′UTR m⁶A modification mediates cap-independent translation. **a**, MEF cells transfected with Fluc mRNA reporters were subject to heat shock treatment and the Fluc activity was measured by real-time luminometry. Fluc activities were quantified and normalized to the sample containing normal adenosine nucleotides. Red, m⁶A; green, 2′-O-MeA. **b**, Constructs expressing Fluc reporter with Hsp70 5′UTR or the one with A103C mutation are depicted on the top. Fluc activities in transfected MEF cells were quantified and normalized to the control containing normal A without stress. **c**, Fluc mRNAs bearing Hsp70 5′UTR with a single m⁶A site were constructed using sequential splint ligation. After *in vitro* translation in rabbit reticulate lysates, Fluc activities were quantified and normalized to the control lacking m⁶A. Error bars, mean ± s.e.m.; *P < 0.05, unpaired two-tailed *t*-test; n = 3 biological replicates (**a**, **b** and **c**). **d**, A proposed model for dynamic m⁶A 5′UTR methylation in response to stress and its role in cap-independent translation. Under the normal growth condition, nuclear FTO demethylates the 5′UTR m⁶A from nascent transcripts and the matured transcripts are translated via a cap-dependent mechanism. Under stress conditions, nuclear localization of YTHDF2 protects the 5′UTR of stress-induced transcripts from demethylation. With enhanced 5′UTR methylation, these transcripts are selectively translated via a cap-independent mechanism.

Hsp105 enhanced translation of the non-capped message after m⁶A incorporation (Extended Data Fig. 9c). This result is consistent with the selective 5′UTR methylation of stress-inducible transcripts upon heat shock stress.

The 5′UTR contains multiple As, although not all of them are methylated. On the basis of the predicted m⁶A sequence motif, the A residue at the 103 position of Hsp70 mRNA is likely to be methylated. Using a single-nucleotide m⁶A detection method[25], we confirmed the methylation event at this position upon heat shock stress (Extended Data Fig. 10a). To demonstrate the significance of methylation at this single site, we introduced an A103C mutation into the Hsp70 5′UTR. Remarkably, m⁶A incorporation no longer promoted translation of the Fluc reporter in transfected cells (Fig. 4b). To directly demonstrate the importance of this single m⁶A site without changing the nucleotide, we employed a sequential RNA splint ligation strategy to construct a Fluc reporter bearing Hsp70 5′UTR with or without A103 methylation (Fig. 4c)[26,27]. Using an *in vitro* translation system, the Fluc reporter containing the single m⁶A at the 103 position showed about 50% increase in translation efficiency in comparison to the one with normal A (Fig. 4c). Notably, both messages showed comparable turnover during the entire course of *in vitro* translation (Extended Data Fig. 10c). Collectively, these results firmly established a crucial role of 5′UTR m⁶A modification in non-canonical translation initiation.

Much of our current understanding of cap-independent translation is limited to the IRES mechanism[28,29]. However, beyond a few examples, many cellular genes capable of cap-independent translation do not seem to contain any IRES elements. The results presented here demonstrate a surprising role of m⁶A in mediating mRNA translation initiation independent of the normal m⁷G cap. How exactly the

methylated adenosine recruits the translation machinery merits further investigation. m6A modification has been shown to alter RNA secondary structures[4]. It is possible that distinct translation initiation factors are recruited to the methylated 5′UTR, thereby facilitating cap-independent translation.

In contrast to the wide belief that m6A modification is static on mRNAs, we found that 5′UTR methylation in the form of m6A is dynamic. Methylation often serves as a mark to distinguish self and foreign DNAs or parental and daughter DNA strands[30]. The stress-inducible mRNA 5′UTR methylation permits ribosomes to distinguish nascent transcripts from pre-existing messages, thereby achieving selective mRNA translation (Fig. 4d). The unexpected stress-inducible feature of YTHDF2 offers an elegant mechanism for temporal control of m6A modification on subsets of mRNAs. The mechanistic connection between 5′UTR methylation and cap-independent translation solves the central puzzle how selective translation is achieved when global translation is suppressed in responding to stress.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N^6-methyladenosine and gene expression control. *Nature Rev. Mol. Cell Biol.* **15,** 313–326 (2014).
2. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m6A RNA methylation. *Nature Rev. Genet.* **15,** 293–306 (2014).
3. Wang, X. et al. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2014).
4. Liu, N. et al. N^6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518,** 560–564 (2015).
5. Wang, X. et al. N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161,** 1388–1399 (2015).
6. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149,** 1635–1646 (2012).
7. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
8. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chem. Biol.* **7,** 885–887 (2011).
9. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49,** 18–29 (2013).
10. Qian, S. B., McDonough, H., Boellmann, F., Cyr, D. M. & Patterson, C. CHIP-mediated stress recovery by sequential ubiquitination of substrates and Hsp70. *Nature* **440,** 551–555 (2006).
11. Schwartz, S. et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. *Cell Rep.* **8,** 284–296 (2014).
12. Schwartz, S. et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155,** 1409–1421 (2013).
13. Anckar, J. & Sistonen, L. Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu. Rev. Biochem.* **80,** 1089–1115 (2011).
14. Mendillo, M. L. et al. HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* **150,** 549–562 (2012).
15. Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Rev. Mol. Cell Biol.* **11,** 113–127 (2010).
16. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83,** 779–812 (2014).
17. Spriggs, K. A., Bushell, M. & Willis, A. E. Translational regulation of gene expression during conditions of cell stress. *Mol. Cell* **40,** 228–237 (2010).
18. Panniers, R. Translational control during heat shock. *Biochimie* **76,** 737–747 (1994).
19. Richter, K., Haslbeck, M. & Buchner, J. The heat shock response: life on the verge of death. *Mol. Cell* **40,** 253–266 (2010).
20. McGarry, T. J. & Lindquist, S. The preferential translation of *Drosophila* hsp70 mRNA requires sequences in the untranslated leader. *Cell* **42,** 903–911 (1985).
21. Klemenz, R., Hultmark, D. & Gehring, W. J. Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. *EMBO J.* **4,** 2053–2060 (1985).
22. Rubtsova, M. P. et al. Distinctive properties of the 5′-untranslated region of human hsp70 mRNA. *J. Biol. Chem.* **278,** 22350–22356 (2003).
23. Sun, J., Conn, C. S., Han, Y., Yeung, V. & Qian, S. B. PI3K–mTORC1 attenuates stress response by inhibiting cap-independent Hsp70 translation. *J. Biol. Chem.* **286,** 6791–6800 (2011).
24. Zhang, X. et al. Translational control of the cytosolic stress response by mitochondrial ribosomal protein L18. *Nature Struct. Mol. Biol.* **22,** 404–410 (2015).
25. Harcourt, E. M., Ehrenschwender, T., Batista, P. J., Chang, H. Y. & Kool, E. T. Identification of a selective polymerase enables detection of N^6-methyladenosine in RNA. *J. Am. Chem. Soc.* **135,** 19079–19082 (2013).
26. Kershaw, C. J. & O'Keefe, R. T. Splint ligation of RNA with T4 DNA ligase. *Methods Mol. Biol.* **941,** 257–269 (2012).
27. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12,** 2014–2019 (2006).
28. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334,** 320–325 (1988).
29. Hellen, C. U. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* **15,** 1593–1612 (2001).
30. Kunkel, T. A. & Erie, D. A. DNA mismatch repair. *Annu. Rev. Biochem.* **74,** 681–710 (2005).

Author Contributions J.Z. and S.-B.Q. conceived the project. J.Z. performed most experiments. J.W. analysed the sequencing data. X.G. performed Ribo-seq. X.Z. assisted heat shock assays. S.R.J. helped with original FTO ideas. S.-B.Q. wrote the manuscript. All authors discussed the results and edited the manuscript.

Author Information Sequencing data have been deposited at NCBI Sequence Read Archive under accession number SRA280261. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.-B.Q. (sq38@cornell.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and reagents. HeLa (cervical cancer) was originally purchased from ATCC and MEF cells were a gift from D. J. Kwiatkowski (Harvard Medical School). Cells were not authenticated recently but tested negative for mycoplasma contamination. Both cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (FBS). Antibodies used in the experiments are listed below: anti-YTHDF2 (Proteintech 24744-1-AP, 1:1,000 WB, 1:600 IF); anti-Hsp70 (Stressgen SPA-810, 1:1,000 WB); anti-FTO (Phosphosolutions 597-Fto, 1:1,000 WB, 1:600 IF); anti-METTL3 (Abnova H00056339-B01P, 1:1,000 WB, 1:600 IF); anti-METTL14 (sigma HPA038002, 1:1,000 WB, 1:600 IF); anti-WTAP (Santa Cruz sc-374280, 1:1,000 WB, 1:600 IF); anti-m^6A (Millipore ABE572, 1:1,000 m^6A immunoblotting); Alexa Fluor 546 donkey anti-mouse secondary antibody (Invitrogen A10036. 1:600 IF); Alexa Fluor 546 donkey anti-rabbit secondary antibody (Invitrogen A10040, 1:600 IF).

Construction of 5′UTR reporters. The Fluc reporter with Hsp70 5′UTR has been reported previously[23]. For Fluc reporters bearing other 5′UTRs, the following primers were used for 5′UTR cloning: Hsc70 (*HSPA8*) forward, 5′-CCCAA GCTTGGTCTCATTGAACGCGG-3′; reverse, 5′-CGGGATCCCCTTAGACA TGGTTGCTT-3′; Tubulin (*TUBG2*) forward, 5′-GGCAAGCTTTGCGCCTGT GCTGAATTCCAGCTGC-3′; reverse, 5′-GGCGGATCCGCATCGCCGATCA GACCTAG-3′; Hsp105 (*HSPH1*) forward, 5′-CCCAAGCTTGTAAAATGCTG CAGATTC-3′; reverse, 5′-CGGGATCCCCACCGACATGGCTGGCCCG-3′.

Lentiviral shRNAs. All shRNA targeting sequences were cloned into DECIPHER pRSI9-U6-(sh)-UbiC-TagRFP-2A-Puro (Cellecta). shRNA targeting sequences listed below were based on RNAi consortium at Broad Institute (http://www.broad.mit.edu/rnai/trc). YTHDF2 (mouse): 5′-GCTCCAGGCATGAATA CTATA-3′; FTO (mouse): 5′-GCTGAGGCAGTTCTGGTTTCA-3′; Scramble control sequence: 5′-AACAGTCGCGTTTGCGACTGG-3′. Lentiviral particles were packaged using Lenti-X 293T cells (Clontech). Virus-containing supernatants were collected at 48 h after transfection and filtered to eliminate cells. MEF cells were infected by the lentivirus for 48 h before selection by 1 μg ml^{-1} puromycin.

Recombinant protein expression. YTHDF2 and FTO were cloned into vector pGEX-6P-1 using the following primers: YTHDF2 forward, 5′-ATGAATTCCC ATCGGCCAGCAGCCTCTTG-3′; reverse, 5′-CCGCTCGAGTTCTATTTCCC ACGACCTTGA-3′; FTO forward, 5′-ATGAATTCAGCATGAAGCGCGTCC AGACC-3′; reverse, 5′-CCGCTCGAGCCTCTAGGATCTTGC-3′.

The resulting clones were transfected into the *Escherichia coli* strain BL21 and expression was induced at 22 °C with 1 mM IPTG for 16–18 h. The pellet collected from 1 l of bacteria culture was then lysed in 15 ml PBS (50 mM NaH_2PO_4, 150 mM NaCl, pH 7.2, 1 mM PMSF, 1 mM DTT, 1 mM EDTA, 0.1% (v/v) Triton X-100) and sonicated for 10 min. After removing cell debris by centrifugation at 12,000 r.p.m. for 30 min, the protein extract was mixed with 2 ml equilibrated Pierce glutathione agarose and mixed on an end-over-end rotator for 2 h at 4 °C. The resin was washed three times with ten resin-bed volumes of equilibration/wash buffer (50 mM Tris, 150 mM NaCl, pH 8.0). YTHDF2 and FTO protein was cleaved from the glutathione agarose using PreScission Protease (Genscript) in cleavage buffer (50 mM Tris-HCl, pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) at 4 °C overnight.

Immunoblotting. Cells were lysed on ice in TBS buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA) containing protease inhibitor cocktail tablet, 1% Triton X-100, and 2 U ml^{-1} DNase. After incubating on ice for 30 min, the lysates were heated for 10 min in SDS/PAGE sample buffer (50 mM Tris (pH 6.8), 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol). Proteins were separated on SDS–PAGE and transferred to Immobilon-P membranes (Millipore). Membranes were blocked for 1 h in TBS containing 5% non-fat milk and 0.1% Tween-20, followed by incubation with primary antibodies overnight at 4 °C. After incubation with horseradish-peroxidase-coupled secondary antibodies at room temperature for 1 h, immunoblots were visualized using enhanced chemiluminescence (ECL^{Plus}, GE Healthcare).

Immunofluorescence staining. Cells grown on glass coverslips were fixed in 4% paraformaldehyde for 10 min at 4 °C. After permeabilization in 0.2% Triton X-100 for 5 min at room temperature, the cover slips were blocked with 1% BSA for 1 h. Cells were stained with indicated primary antibody overnight at 4 °C, followed by incubation with Alexa Fluor 546 donkey anti-mouse secondary antibody or Alexa Fluor 546 donkey anti-rabbit secondary antibody for 1 h at room temperature. The nuclei were counter-stained with DAPI (1:1,000 dilution) for 10 min. Cover slips were mounted onto slides and visualized using a Zeiss LSM710 confocal microscope.

mRNA stability measurement. Cells were treated with actinomycin D (5 μg ml^{-1}) for 4 h, 2 h and 0 h before trypsinization and collection. RNA spike-in control was added proportional to the total cell numbers and total RNA was isolated by TRIzol kit (Life Technologies). After reverse transcription, the mRNA levels of transcripts of interest were detected by real-time quantitative PCR.

Real-time quantitative PCR. Total RNA was isolated by TRIzol reagent (Invitrogen) and reverse transcription was performed using High Capacity cDNA Reverse Transcription Kit (Invitrogen). Real-time PCR analysis was conducted using Power SYBR Green PCR Master Mix (Applied Biosystems) and carried on a LightCycler 480 Real-Time PCR System (Roche Applied Science). Primers for amplifying each target were: YTHDF2 forward, 5′-CAGTTTGCCT CCAGCTACTATT-3′; reverse, 5′-GCAATGCCATTCTTGGTCTTC-3′; FTO forward, 5′-TCAGCAGTGGCAGCTGAAAT-3′; reverse, 5′-CTTGGATCCTC ACCACGTCC-3′; Hsp70 forward, 5′-TGGTGCAGTCCGACATGAAG-3′; reverse, 5′-GCTGAGAGTCGTTGAAGTAGGC-3′; METTL3 forward, 5′-ATC CAGGCCCATAAGAAACAG-3′; reverse, 5′-CTATCACTACGGAAGGTTG GG-3′; METTL14 forward, 5′-CAGGCAGGACATGGGGATATT-3′; reverse, 5′-TCCGACCTGGAGACATACAT-3′; ALKBH5 forward, 5′-AGTTCCAGTTC AAGCCCATC-3′; reverse, 5′-GGCGTTCCTTAATGTCCTGAG-3′; WTAP forward, 5′-CTGGCAGAGGAGGTAGTAGTTA-3′; reverse, 5′-ACTGGAGTCTG TGTCATTTGAG-3′; β-actin forward, 5′-TTGCTGACAGGATGCAGAAG-3′; reverse, 5′-ACTCCTGCTTGCTGATCCACAT-3′; GAPDH forward, 5′-CAAG GAGTAAGAAACCCTGGAC-3′; reverse, 5′-GGATGGAAATTGTGAGGGAG AT-3′; Fluc forward, 5′-ATCCGGAAGCGACCAACGCC-3′; reverse, 5′-GTCG GGAAGACCTGCCACGC-3′.

In vitro transcription. Plasmids containing the corresponding 5′UTR sequences of mouse *HSPA1A* and full-length firefly luciferase were used as templates. Transcripts with normal m^7G cap were generated using the mMessage mMachine T7 Ultra kit (Ambion) and transcripts with non-functional cap analogue GpppA were synthesized using MEGAscript T7 Transcription Kit (Ambion). To obtain mRNAs with the adenosine replaced with m^6A, *in vitro* transcription was conducted in a reaction in which 5% of the adenosine was replaced with N^6-methyladenosine. All mRNA products were purified using the MEGAclear kit (Ambion) according to the manufacturer's instructions.

In vitro translation. *In vitro* translation was performed using the Rabbit Reticulocyte Lysate System (Promega) according to the manufacturer's instructions. Luciferase activity was measured using a luciferase reporter assay system (Promega) on a Synergy HT Multi-detection Microplate Reader (BioTek Instruments).

Real-time luciferase assay. Cells grown in 35-mm dishes were transfected with *in-vitro*-synthesized mRNA containing the luciferase gene. Luciferase substrate D-luciferin (1 mM, Regis Tech) was added into the culture medium immediately after transfection. Luciferase activity was monitored and recorded using Kronos Dio Luminometer (Atto).

Site-specific m^6A detection. For site-specific m^6A detection, DNA primers were first 5′ labelled with ^{32}P using T4 polynucleotide kinase (Invitrogen) and purified by ethanol precipitation. The primer 5′-AGGGATGCTCTGGGGAAGGCTGG-3′ was used to detect potential m^6A site and the primer 5′-CGCCGCTCG CTCTGCTTCTCTTGTCTTCGCT-3′ was used to detect the non-methylated site. Synthesized mRNA 5′-CGATCCTCGGCCAGG(**m^{6}A**)CCAGCCTTCCCC AG-3′ and 5′-CGATCCTCGGCCAGG**A**CCAGCCTTCCCCAG-3′ served as positive and negative control templates, respectively. To set up the reaction, a $2 \times$ annealing solution was prepared in a total volume of 8 μl with $1 \times$ Tth buffer (Promega) or AMV buffer (Invitrogen), 1 μl of each radiolabelled primer and 10 μg mRNA from MEF cells that had been heat shock treated. The mixture was heated at 95 °C for 10 min and cooled slowly to room temperature. 3 μl of annealing solution were combined with 2 μl of enzyme and heated at 37 °C (AMV Reverse Transcriptase) or 55 °C (Tth DNA Polymerase) for 2 min. After adding the dTTP solution (final dTTP concentration: 100 μM), the reactions were heated for 5 min at 37 °C (AMV) or 10 min at 55 °C (Tth). Reaction products were resolved on a 20% denaturing polyacrylamide gel and exposed overnight.

RNA splint ligation. The ligation method was optimized from previous reports[26,27,31]. The RNA oligonucleotide covering the 82–117 nt region of *HSPA1A* was synthesized by Thermo Scientific, whereas RNA fragments corresponding to other regions were generated by *in vitro* transcription. For sequential splint ligation, two DNA bridging oligonucleotides were designed: DNA Bridge 1, 5′-GGTCCTGGCCGAGGATCGGGAACGCGCCGCTCGCTC-3′; DNA Bridge 2, 5′-CTCCGCGGCAGGGATGCTCTGGGGAAGGCTGGTCCT-3′.

For 3′ RNA oligonucleotide (donor) phosphorylation, 1 μl of 20 μM donor oligonucleotide was mixed with 1 μl of 10× PNK buffer, 6 μl of ATP (10 mM), 0.5 μl of RNasin (20 units) and 1 μl of T4 PNK (5 units). The reaction mixture was incubated at 37 °C for 30 min followed by inactivation of T4 PNK at 65 °C for 20 min. Next, the DNA bridge oligonucleotide was hybridized with the 3′ RNA

oligonucleotide and the 5′ RNA oligonucleotide (acceptor) at a 1:1.5:2 ratio (5′RNA:bridge:3′RNA). Oligonucleotides were annealed (95 °C for 1 min followed by 65 °C for 2 min and 37 °C for 10 min) in the presence of $1 \times$ T4 DNA dilution buffer. To ligate the 5′ and the 3′ RNA together, T4 DNA ligase and the T4 DNA ligation buffer were added and the reaction mixture was incubated at 37 °C for 1 h. The ligation was stopped by adding 1 μl of 0.5 M EDTA followed by phenol-chloroform extraction and ethanol precipitation. Ligation products were analysed by 10% TBE-Urea gels or formaldehyde gels. The expected RNA ligation products in TBE-Urea gels were eluted in RNA gel elution buffer (300 mM NaOAc pH 5.5, 1 mM EDTA and $0.1 \, \text{U} \, \mu\text{l}^{-1}$ SUPERase_In) followed by ethanol precipitation. The final products in formaldehyde gels were isolated by Zymoclean Gel RNA Recovery Kit (Zymo Research).

Hsp70 mRNA pull-down and m^{6}A immunoblotting. To isolate endogenous *Hsp70* mRNA, 400 pmol of biotin-labelled probe (5′-TTCATAACATATCTCT GTCTCTT-3′) was incubated with 2 mg M-280 Streptavidin Dynabeads (Life Technologies) in 1 ml $1 \times$ B & W buffer (5 mM Tris-HCL pH 7.5, 0.5 mM EDTA and 1 M NaCl) at 4 °C for 1 h. 2 mg total RNA was denatured at 75 °C for 2 min and added to the pre-coated Dynabeads for an additional incubation of 2 h at 4 °C. Captured RNA was eluted by heating beads for 2 min at 90 °C in 10 mM EDTA with 95% formamide followed by TRIzol LS isolation. Isolated RNA was quantified using NanoDrop ND-1000 UV-Vis Spectrophotometer and equal amounts of RNAs were mixed with $2 \times$ RNA Loading Dye (Thermo Scientific) and denatured for 3 min at 70 °C. *In-vitro*-transcribed mRNA containing 50% N^6-methyl-ladenosine or 100% adenosine was used as positive and negative control, respectively. Samples were separated on a formaldehyde denaturing agarose gel and transferred to a positively charged nylon membrane by siphonage in transfer buffer (10 mM NaOH, 3 M NaCl) overnight at room temperature. After transfer, the membrane was washed for 5 min in $2 \times$ SSC buffer and RNA was UV cross-linked to the membrane. Membrane was blocked for 1 h in PBST containing 5% non-fat milk and 0.1% Tween-20, followed by incubation with anti-m^{6}A antibody (1:1,000 dilution) for overnight at 4 °C. After extensive washing with 0.1% PBST three times, the membrane was incubated with HRP-conjugated anti-rabbit IgG (1:5,000 dilution) for 1 h. Membrane was visualized by using enhanced chemiluminescence (ECLPlus, GE Healthcare).

YTHDF2 and FTO *in vitro* pull down. Synthesize mRNA (100 pmol) with a single m^{6}A at A103 was label by biotin using the Pierce RNA 3′ End Desthiobiotinylation Kit. Binding of the labelled RNA to streptavidin magnetic beads was performed in RNA capture buffer (20 mM Tris, pH 7.5, 1 M NaCl, 1 mM EDTA) for 30 min at room temperature with rotation. The RNA–protein binding reaction was conducted in protein–RNA binding buffer (20 mM Tris (pH 7.5), 50 M NaCl, 2 mM MgCl$_2$, 0.1% Tween-20 Detergent) at 4 °C for 60 min with rotation. After washing three times with the wash buffer (20 mM Tris pH 7.5, 10 mM NaCl, 0.1% Tween-20 Detergent), protein was eluted by Biotin Elution Buffer (Pierce) and detected by western blot.

YTHDF2 and FTO *in vitro* competition assay. The YTHDF2 and FTO *in vitro* competition assay was performed in 100 μl of reaction mixture containing 5 μM RNA incorporated with 50% m^{6}A, 283 μM of $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, 300 μM of α-KG, 2 mM of L-ascorbic acid, 50 μg ml^{-1} of BSA, and 50 mM of HEPES buffer (pH 7.0). The reaction was incubated for 3 h at room temperature, and quenched by adding 5 mM EDTA followed by heating for 5 min at 95 °C. RNA was isolated by TRIzol LS and quantified using NanoDrop ND-1000 UV-Vis Spectrophotometer. Equal amounts of RNA were used for dot blotting and methylene blue staining was used to show the amount of RNA on hybridization membranes.

Polysome profiling analysis. Sucrose solutions were prepared in polysome buffer (10 mM HEPES, pH 7.4, 100 mM KCl, 5 mM MgCl$_2$, 100 μg ml^{-1} cycloheximide and 2% Triton X-100). A 15%–45% (w/v) Sucrose density gradients were freshly prepared in SW41 ultracentrifuge tubes (Backman) using a Gradient Master (BioComp Instruments). Cells were pre-treated with 100 μg ml^{-1} cycloheximide for 3 min at 37 °C followed by washing using ice-cold PBS containing 100 μg ml^{-1} cycloheximide. Cells were then lysed in polysome lysis buffer. Cell debris were removed by centrifugation at 14,000 r.p.m. for 10 min at 4 °C. 500 μl of supernatant was loaded onto sucrose gradients followed by centrifugation for 2 h 28 min at 38,000 r.p.m. 4 °C in a SW41 rotor. Separated samples were fractionated at 0.75 ml min^{-1} through an automated fractionation system (Isco) that continually monitores OD$_{254}$ values. An aliquot of ribosome fraction were used to extract total RNA using Trizol LS reagent (Invitrogen) for real-time PCR analysis.

RNA-seq and m^{6}A-seq. For m^{6}A immunoprecipitation, total RNA was first isolated using TRIzol reagent followed by fragmentation using freshly prepared RNA fragmentation buffer (10 mM Tris-HCl, pH 7.0, 10 mM ZnCl$_2$). 5 μg fragmented RNA was saved as input control for RNA-seq. 1 mg fragmented RNA was incubated with 15 μg anti-m^{6}A antibody (Millipore ABE572) in $1 \times$ IP buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 0.1% Igepal CA-630) for 2 h at 4 °C. The m^{6}A-IP mixture was then incubated with Protein A beads for additional 2 h at 4 °C

on a rotating wheel. After washing three times with IP buffer, bound RNA was eluted using 100 μl elution buffer (6.7 mM N^6-Methyladenosine 5′-monophosphate sodium salt in $1 \times$ IP buffer), followed by ethanol precipitation. Precipitated RNA was used for cDNA library construction and high-throughput sequencing described below.

Ribo-seq. Ribosome fractions separated by sucrose gradient sedimentation were pooled and digested with *E. coli* RNase I (Ambion, 750 U per 100 A260 units) by incubation at 4 °C for 1 h. SUPERase inhibitor (50 U per 100 U RNase I) was then added into the reaction mixture to stop digestion. Total RNA was extracted using TRIzol reagent. Purified RNA was used for cDNA library construction and high-throughput sequencing described below.

cDNA library construction. Fragmented RNA input and m^{6}A-IP elutes were dephosphorylated for 1 h at 37 °C in 15 μl reaction ($1 \times$ T4 polynucleotide kinase buffer, 10 U SUPERase_In and 20 U T4 polynucleotide kinase). The products were separated on a 15% polyacrylamide TBE-urea gel (Invitrogen) and visualized using SYBR Gold (Invitrogen). Selected regions of the gel corresponding to 40–60 nt (for RNA-seq and m^{6}A-seq) or 25–35 nt (for Ribo-seq) were excised. The gel slices were disrupted by using centrifugation through the holes at the bottom of the tube. RNA fragments were dissolved by soaking overnight in 400 μl gel elution buffer (300 mM NaOAc, pH 5.5, 1 mM EDTA, $0.1 \, \text{U} \, \mu\text{l}^{-1}$ SUPERase_In). The gel debris was removed using a Spin-X column (Corning), followed by ethanol precipitation. Purified RNA fragments were re-suspended in nuclease-free water. Poly(A) tailing reaction was carried out for 45 min at 37 °C ($1 \times$ poly(A) polymerase buffer, 1 mM ATP, $0.75 \, \text{U} \, \mu\text{l}^{-1}$ SUPERase_ In and 3 U *E. coli* poly(A) polymerase).

For reverse transcription, the following oligonucleotides containing barcodes were used: MCA02, 5′-pCAGATCGTCGGACTGTAGAACTCTØCAAGCAGA AGACGGCATACGATTTTTTTTTTTTTTTTTTTTTTTVN-3′; LGT03, 5′-pGTG ATCGTCGGACTGTAGAACTCTØCAAGCAGAAGACGGCATACGATTTT TTTTTTTTTTTTTTTVN-3′; YAG04, 5′-pAGGATCGTCGGACTGTAGA ACTCTØCAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTTTTV-N-3′; HTC05, 5′-pTCGATCGTCGGACTGTAGAACTCTØCAAGCAGAAGA CGGCATACGATTTTTTTTTTTTTTTTTTTTTVN-3′.

In brief, the tailed-RNA sample was mixed with 0.5 mM dNTP and 2.5 mM synthesized primer and incubated at 65 °C for 5 min, followed by incubation on ice for 5 min. The following was then added to the reaction mix: 20 mM Tris (pH 8.4), 50 mM KCl, 5 mM MgCl$_2$, 10 mM DTT, 40 U RNaseOUT and 200 U SuperScript III. The reverse-transcription reaction was performed according to the manufacturer's instruction. Reverse-transcription products were separated on a 10% polyacrylamide TBE-urea gel as described earlier. The extended first-strand product band was expected to be approximately 100 nt, and the corresponding region was excised. The cDNA was recovered by using DNA gel elution buffer (300 mM NaCl, 1 mM EDTA). First-strand cDNA was circularized in 20 μl of reaction containing $1 \times$ CircLigase buffer, 2.5 mM MnCl$_2$, 1 M Betaine, and 100 U CircLigase II (Epicentre). Circularization was performed at 60 °C for 1 h, and the reaction was heat-inactivated at 80 °C for 10 min. Circular single-strand DNA was re-linearized with 20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, and 7.5 U APE 1 (NEB). The reaction was carried out at 37 °C for 1 h. The linearized single-strand DNA was separated on a Novex 10% polyacrylamide TBE-urea gel (Invitrogen) as described earlier. The expected 100-nt product bands were excised and recovered as described earlier.

Deep sequencing. Single-stranded template was amplified by PCR by using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions. The oligonucleotide primers qNTI200 (5′-CAAGCAGAAGACGGCATA-3′) and qNTI201 (5′-AATGATACGGCGACCACCGACAGGTTCAGAGTTC TACAGTCCGACG- 3′) were used to create DNA suitable for sequencing, that is, DNA with Illumina cluster generation sequences on each end and a sequencing primer binding site. The PCR contains $1 \times$ HF buffer, 0.2 mM dNTP, 0.5 μM oligonucleotide primers, and 0.5 U Phusion polymerase. PCR was carried out with an initial 30 s denaturation at 98 °C, followed by 12 cycles of 10 s denaturation at 98 °C, 20 s annealing at 60 °C, and 10 s extension at 72 °C. PCR products were separated on a non-denaturing 8% polycrylamide TBE gel as described earlier. Expected DNA at 120 bp (for Ribo-seq), or 140 bp (for RNA-seq and m^{6}A-seq) was excised and recovered as described earlier. After quantification by Agilent BioAnalyzer DNA 1000 assay, equal amounts of barcoded samples were pooled into one sample. Approximately 3–5 pM mixed DNA samples were used for cluster generation followed by deep sequencing using sequencing primer 5′-CGACAGGTTCAGAGTTCTAC AGTCCGACGATC-3′ (Illumina HiSeq).

Preprocessing of sequencing reads. For Ribo-seq, the sequencing reads were first trimmed by 8 nt from the 3′ end and trimmed reads were further processed by removing the adenosine (A) stretch from the 3′ end (one mismatch was allowed). The processed reads between 25 nt and 35 nt were first mapped by Tophat using parameters (--bowtie1 -p 10 --no-novel-juncs) to mouse transcriptome (UCSC Genes)[32]. The unmapped reads were then mapped to corresponding mouse

genome (mm10). Non-uniquely mapped reads were disregarded for further analysis owing to ambiguity. The same mapping procedure was applied to RNA-seq and m⁶A-seq. For Ribo-seq, the 13th position (12 nt offset from the 5′ end) of the uniquely mapped read was defined as the ribosome 'P-site' position. The RPF density was computed after mapping uniquely mapped reads to each individual mRNA transcript according to the NCBI Refseq gene annotation. Uniquely mapped reads of RNA-seq and Ribo-seq in the mRNA coding region were used to calculate the RPKM values for estimating mRNA expression and translation levels respectively. For m⁶A-seq, uniquely mapped reads in the 5′UTR were used to calculate the RPKM values for estimating the m⁶A levels.

Identification of m⁶A sites. We used a similar scanning strategy reported previously to identify m⁶A peaks in the immunoprecipitation sample as compared to the input sample[7]. In brief, for NCBI RefSeq genes whose maximal read coverage was greater than 15 in the input (RNA-seq), a sliding window of 80 nucleotides with step size of 40 nucleotides was employed to scan the longest isoform (on the basis of coding sequence (CDS) length; in the case of equal CDS, the isoform with longer 5′UTR was selected). For each window, a peak-over-median score (POM) was derived by calculating the ratio of mean read coverage in the window to the median read coverage of the whole gene body. Windows scoring higher than 3 in

the IP sample were obtained and all the resultant overlapping m⁶A peak windows in the IP sample were iteratively clustered to infer the boundary of the m⁶A-enriched region, as well as peak position with maximal read coverage. Finally, a peak-over-input (POI) score was assigned to each m⁶A-enriched region by calculating the ratio of POM in the IP sample to that in the input sample. A putative m⁶A site was defined if the POI score was higher than 3. The peak position of each m⁶A site was classified into five mutually exclusive mRNA structural regions including TSS (the first 200 nucleotides of mRNA), 5′UTR, CDS, stop codon (a 400 nt window flanking the mRNA stop codon) and 3′UTR.
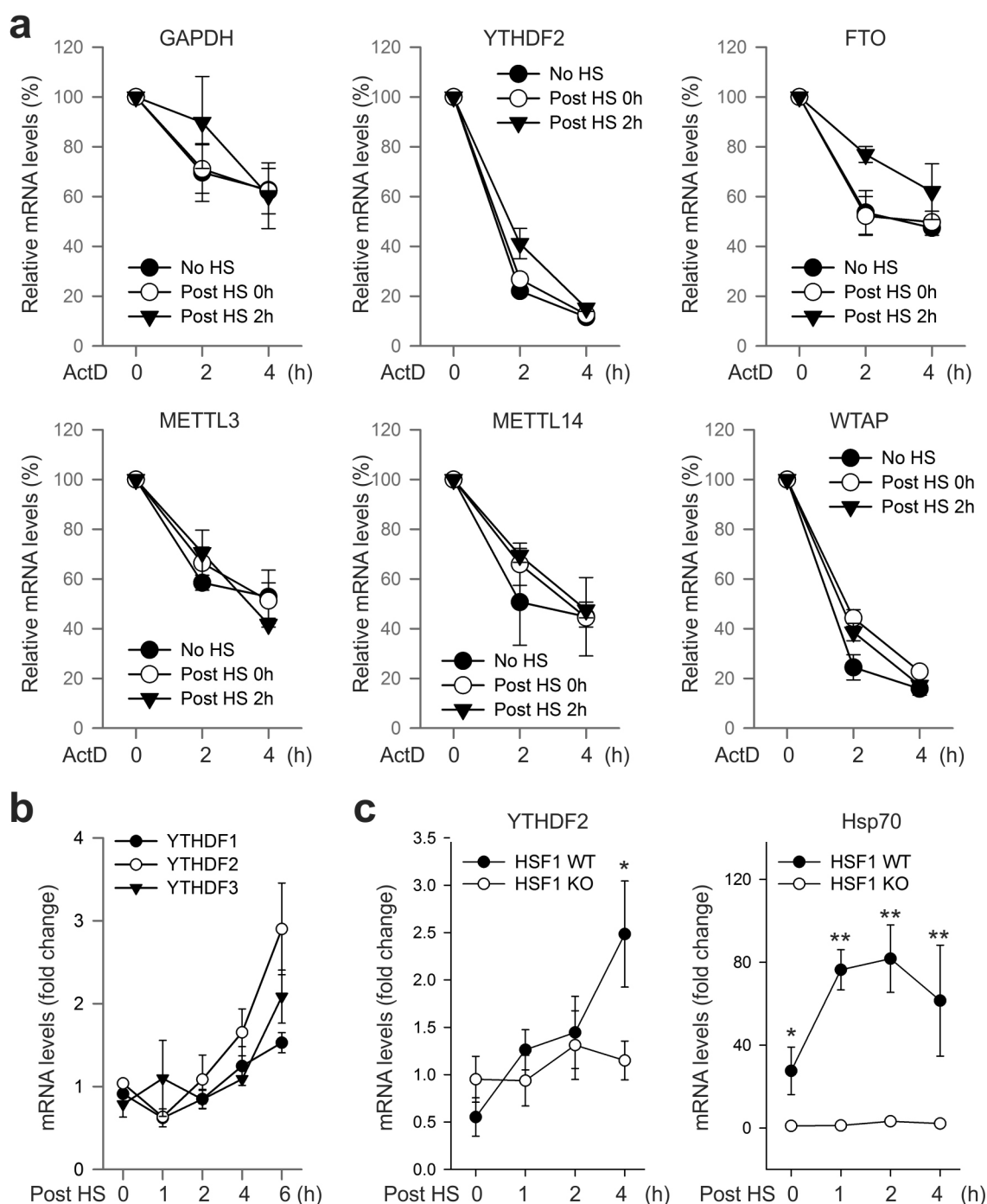
m⁶A motif analysis. The m⁶A peaks with POI score higher than 10 were selected for consensus motif finding. We used MEME Suite for motif analysis[33]. In brief, the flanking sequences of m⁶A peaks (±40 nt) with POI scores were retrieved from mouse transcritpome and were used as MEME input.
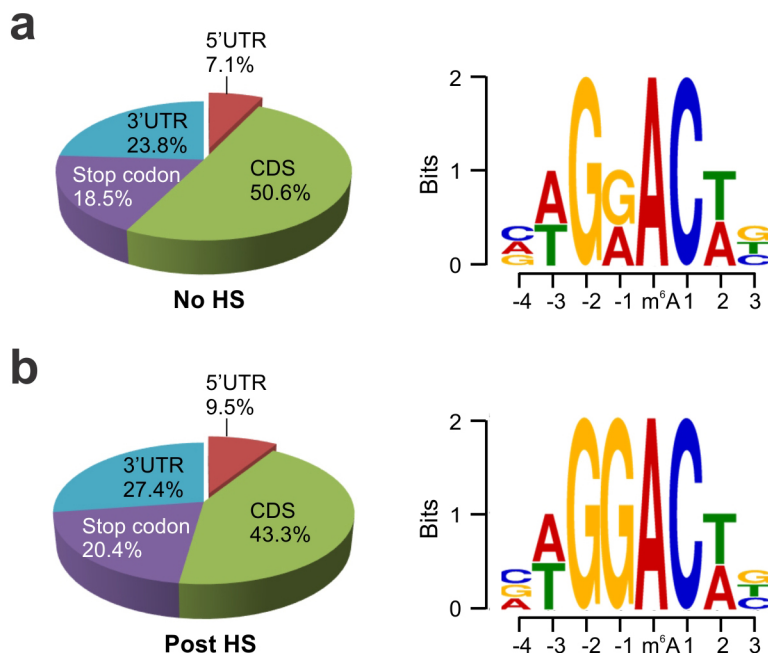
31. Maroney, P. A., Chamnongpol, S., Souret, F. & Nilsen, T. W. Direct detection of small RNAs using splinted ligation. *Nature Protocols* **3,** 279–287 (2008).
32. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25,** 1105–1111 (2009).
33. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).

a

b
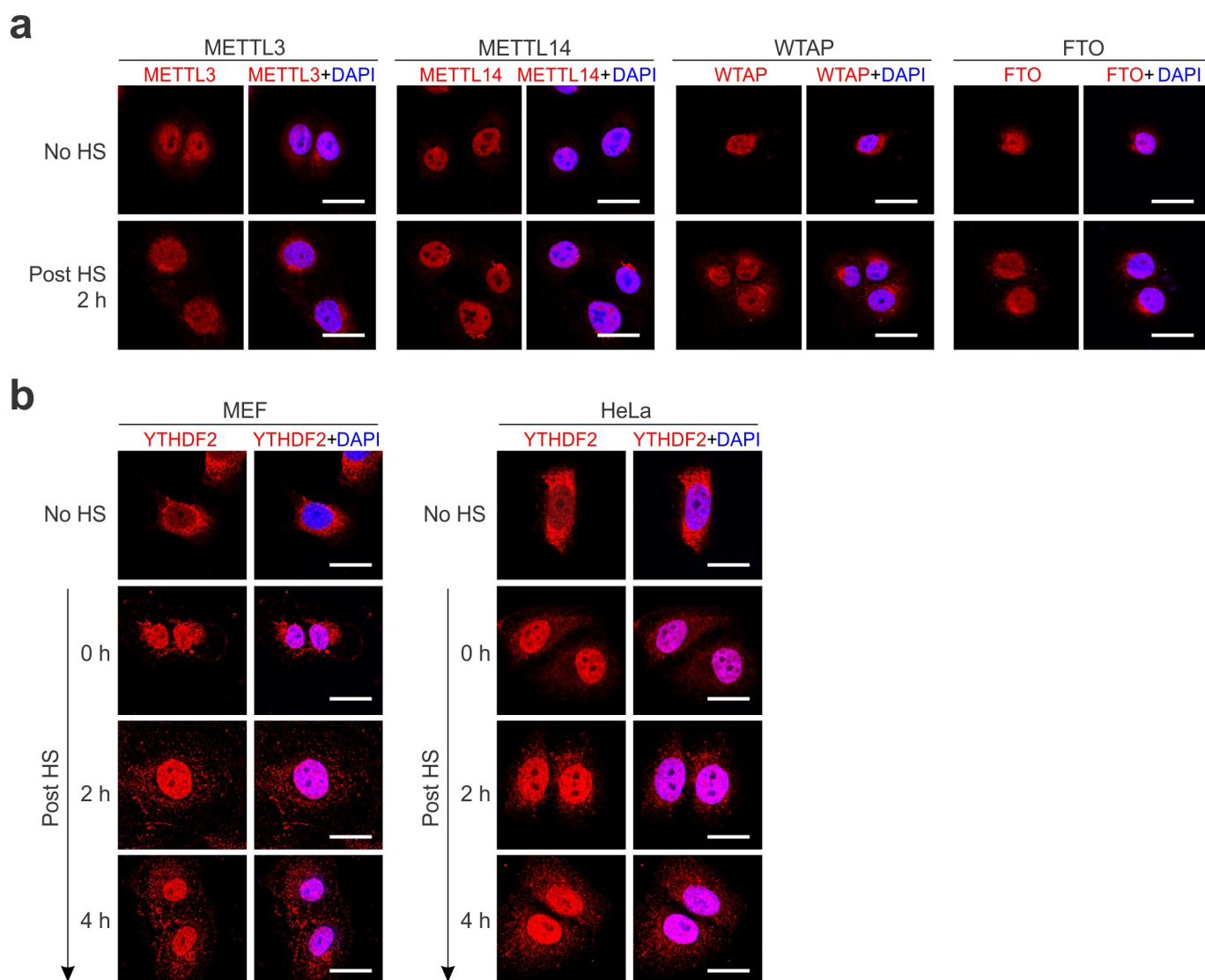
Extended Data Figure 1 | Subcellular localization of the m⁶A machinery in cells before and after heat shock stress. **a**, MEF cells before or 2 h after heat shock (HS; 42 °C, 1 h) were immunostanied with indicated antibodies. DAPI was used for nuclear staining. **b**, MEF (left panel) and HeLa cells (right panel) were subject to heat shock stress (42 °C, 1 h) followed by recovery at 37 °C for various times. Anti-YTHDF2 immunostaining was counterstained by DAPI. Images are representative of at least 50 cells. Bar, 10 μm.

Extended Data Figure 2 | mRNA stability and induction in response to heat shock stress. **a**, Effects of heat shock stress on mRNA stability. MEF cells without heat shock stress (No HS), immediately after heat shock stress (42 °C, 1 h) (Post HS 0 h), or 2 h recovery at 37 °C (Post HS 2 h) were subject to further incubation in the presence of 5 μg ml^{-1} ActD. At the indicated times, mRNA levels were determined by qPCR. Error bars, mean ± s.e.m.; $n = 3$ biological replicates. **b**, MEF cells were collected at indicated times after heat shock stress (42 °C, 1 h) followed by RNA extraction and real-time PCR.

Relative levels of indicated transcripts are normalized to β-actin. Error bars, mean ± s.e.m.; $n = 3$ biological replicates. **c**, HSF1 wild-type (WT) and knockout (KO) cells were subject to heat shock stress (42 °C, 1 h) followed by recovery at 37 °C for various times. Real-time PCR was conducted to quantify transcripts encoding Hsp70 and YTHDF2. Relative levels of transcripts are normalized to β-actin. Error bars, mean ± s.e.m.; *$P < 0.05$, **$P < 0.01$, unpaired two-tailed t-test; $n = 3$ biological replicates.

a

b

Extended Data Figure 3 | Characterization of m^{6}A sites in MEF cells with or without heat shock stress. **a**, **b**, m^{6}A profiling was conducted on MEF cells before (**a**) or 2 h after (**b**) heat shock (42 °C, 1 h). Left, pie chart presenting fractions of m^{6}A peaks in different transcript segments. Right, sequence logo representing the consensus motif relative to m^{6}A. CDS, coding sequence region.
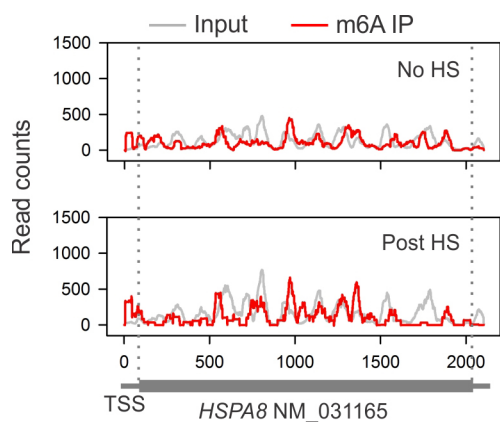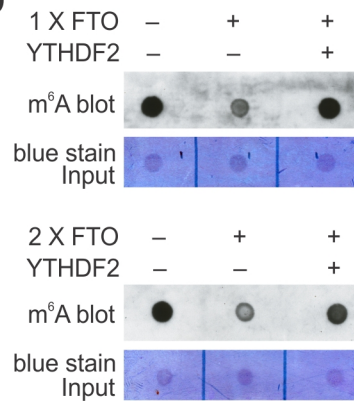
Extended Data Figure 4 | m⁶A profiling of *HSPA8* in MEF cells with or without heat shock stress. An example of constitutively expressed transcript *HSPA8* in MEF cells with or without heat shock stress. Coverage of m⁶A immunoprecipitation and control reads (input) are indicated in red and grey, respectively. The transcript architecture is shown below the *x* axis.
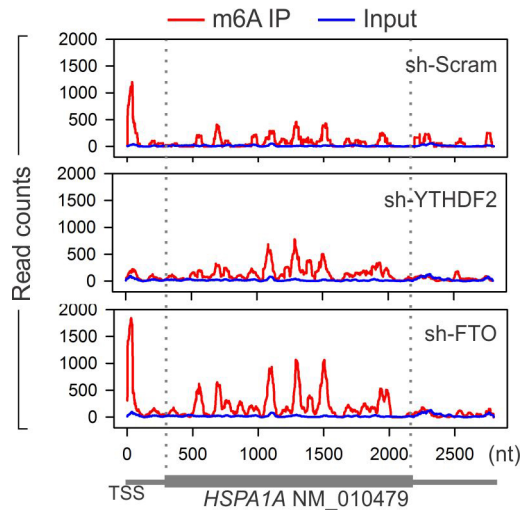
Extended Data Figure 5 | Dynamic m⁶A modification of *HSPA1A* by YTHDF2 and FTO. An example of stress-induced transcript *HSPA1A* in post-stressed MEF cells with either YTHDF2 or FTO knockdown. Coverage of m⁶A immunoprecipitation and control reads (input) are indicated in red and blue, respectively. The transcript architecture is shown below the *x* axis.

a

Hsp70 5'UTR(m⁶A)

b
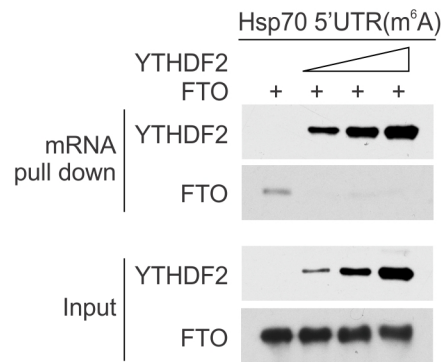
Extended Data Figure 6 | Direct competition between YTHDF2 and FTO in m⁶A binding. a, Synthesized mRNA with m⁶A was incubated with FTO (2 μg) in the presence of an increasing amount of YTHDF2 (0, 0.5, 1 and 2 μg), followed by RNA pull-down and immunoblotting. **b,** Synthesized mRNA with m⁶A was incubated with FTO (1 μg in top panel and 2 μg in bottom panel) in the absence of presence of YTHDF2 (4 μg), followed by m⁶A dot blotting.
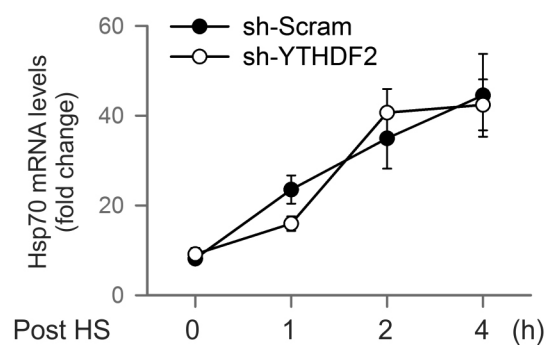
Extended Data Figure 7 | YTHDF2 knockdown does not affect Hsp70 transcription after stress. MEF cells with or without YTHDF2 knockdown were subject to heat shock stress (42 °C, 1 h) followed by recovery at 37 °C for various times. Real-time PCR was conducted to quantify Hsp70 mRNA levels. Error bars, mean ± s.e.m.; $n = 3$ biological replicates. sh-Scram, scrambled shRNA.
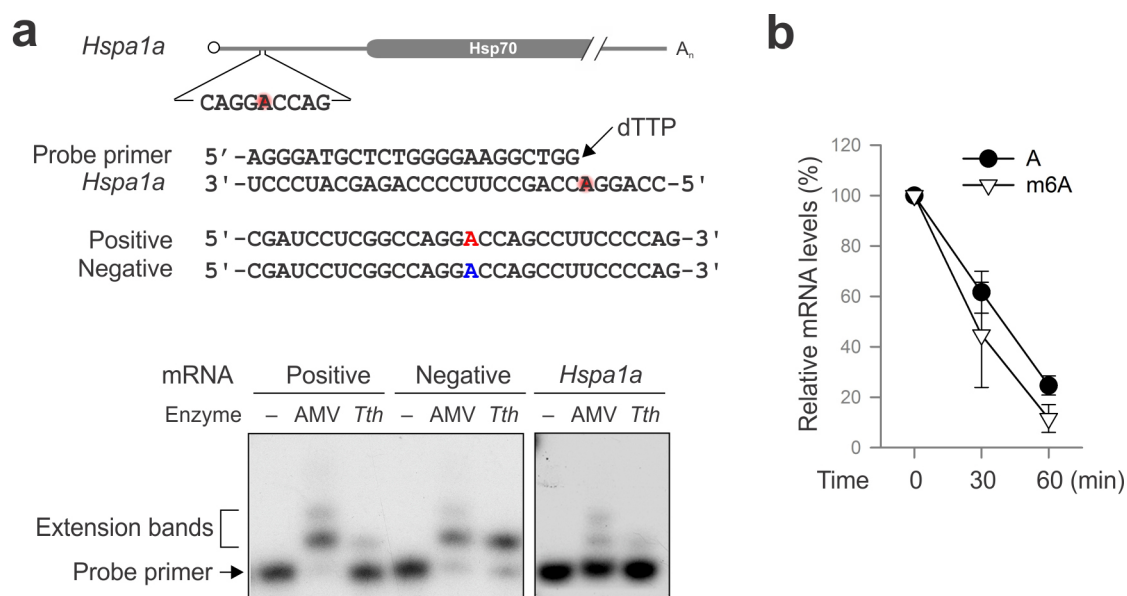
a

RNA staining   m6A blot

b
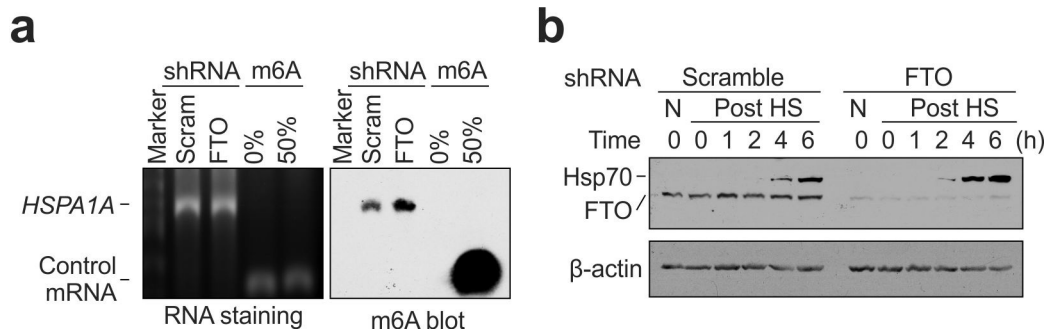
Extended Data Figure 8 | FTO knockdown promotes Hsp70 synthesis.
a, m⁶A blotting of purified *HSPA1A* in MEF with or without FTO knockdown. Messenger RNAs synthesized by *in vitro* transcription in the absence or presence of m⁶A were used as control. RNA staining is shown as loading control. Representative of two biological replicates. **b**, MEF cells with or without FTO knockdown were collected at indicated times after heat shock stress (42 °C, 1 h) followed by immunoblotting using antibodies indicated. N, no heat shock. Representative of three biological replicates.
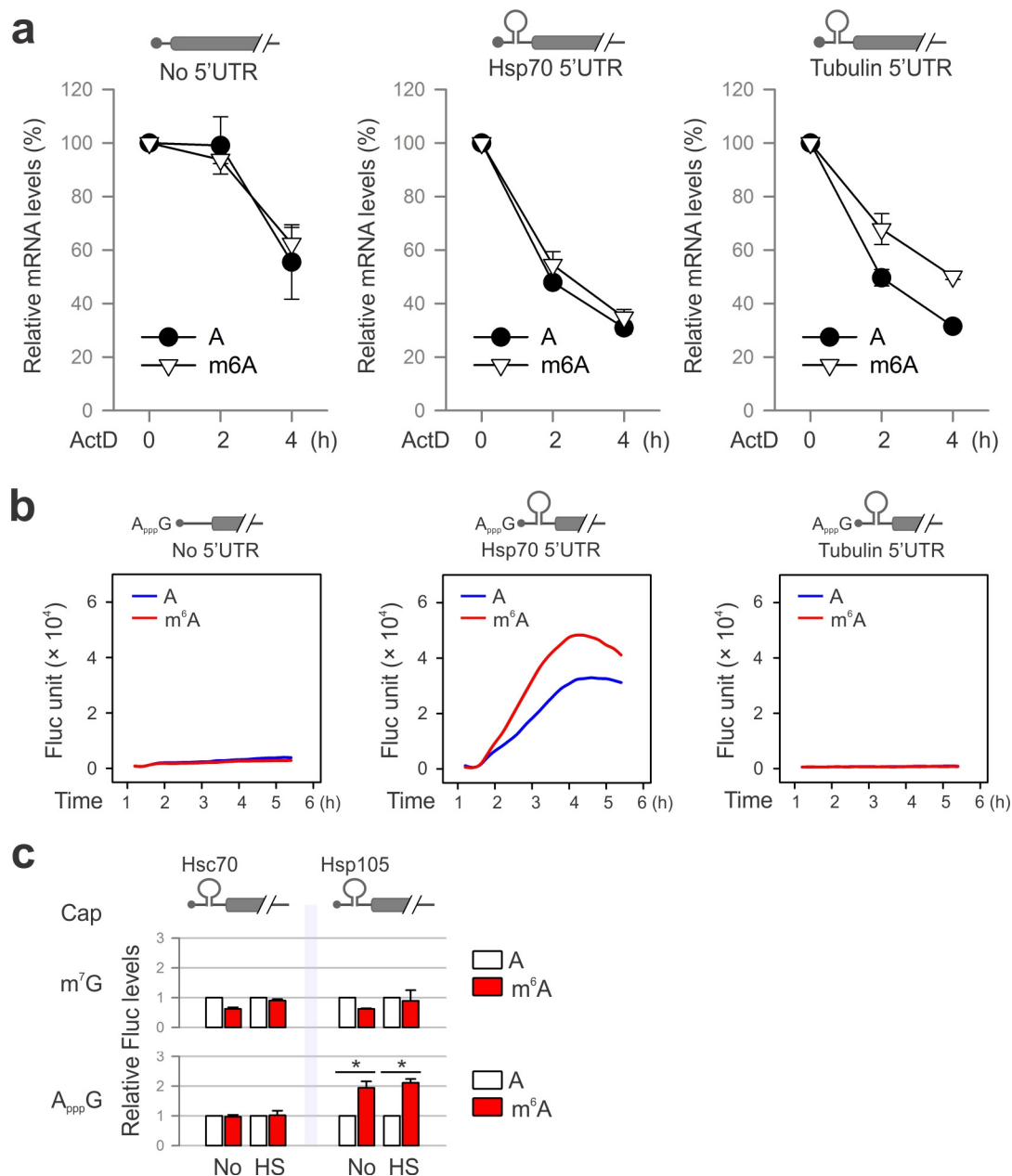
Extended Data Figure 9 | m⁶A modification promotes cap-independent translation. **a**, Fluc reporter mRNAs with or without 5′UTR was synthesized in the absence or presence of m⁶A. The transfected MEFs were incubation in the presence of 5 μg ml⁻¹ ActD. At the indicated times, mRNA levels were determined by qPCR. Error bars, mean ± s.e.m.; $n = 3$ biological replicates. **b**, Fluc reporter mRNAs with or without Hsp70 5′UTR was synthesized in the absence of presence of m⁶A, followed by addition of a non-functional cap analogue ApppG. Fluc activity in transfected MEF cells was recorded using real-time luminometry. **c**, Constructs expressing Fluc reporter bearing 5′UTR from Hsc70 or Hsp105 are depicted on the top. Fluc activities in transfected MEF cells were quantified and normalized to the control containing normal A. Error bars, mean ± s.e.m.; *$P < 0.05$, unpaired two-tailed t-test; $n = 3$ biological replicates.

a

Hspa1a

CAGG**A**CCAG

dTTP

Probe primer 5′–AGGGATGCTCTGGGGAAGGCTGG

Hspa1a 3′–UCCCUACGAGACCCCUUCCGACC**A**GGACC–5′

Positive 5′–CGAUCCUCGGCCAGG**A**CCAGCCUUCCCCAG–3′

Negative 5′–CGAUCCUCGGCCAGG**A**CCAGCCUUCCCCAG–3′

mRNA	Positive			Negative			*Hspa1a*		
Enzyme	–	AMV	*Tth*	–	AMV	*Tth*	–	AMV	*Tth*

Extension bands

Probe primer →

b

Extended Data Figure 10 | Site-specific detection of m⁶A modification on *HSPA1A*. a, Sequences of *HSPA1A* template and the DNA primer used for site-specific detection. Synthesized mRNAs containing a single m⁶A site (red) or A (blue) are used as positive and negative controls, respectively. The red shading in the *HSPA1A* sequence indicates predicted m⁶A sites. Autoradiogram shows primer extension of controls (left panel) and endogenous *HSPA1A* (right panel). **b**, Fluc mRNAs with or without m⁶A incorporation were incubated in the rabbit reticulocyte lysate system (RRL) at 30 °C for up to 60 min. Messenger RNA levels were determined by qPCR. Error bars, mean ± s.e.m.; *n* = 3 biological replicates.

CORRECTIONS & AMENDMENTS

Erratum: Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone

Yi-Gang Tong, Wei-Feng Shi, Di Liu, Jun Qian, Long Liang,
Xiao-Chen Bo, Jun Liu, Hong-Guang Ren, Hang Fan, Ming Ni,
Yang Sun, Yuan Jin, Yue Teng, Zhen Li, David Kargbo,
Foday Dafae, Alex Kanu, Cheng-Chao Chen, Zhi-Heng Lan,
Hui Jiang, Yang Luo, Hui-Jun Lu, Xiao-Guang Zhang, Fan Yang,
Yi Hu, Yu-Xi Cao, Yong-Qiang Deng, Hao-Xiang Su, Yu Sun,
Wen-Sen Liu, Zhuang Wang, Cheng-Yu Wang, Zhao-Yang Bu,
Zhen-Dong Guo, Liu-Bo Zhang, Wei-Min Nie, Chang-Qing Bai,
Chun-Hua Sun, Xiao-Ping An, Pei-Song Xu,
Xiang-Li-Lan Zhang, Yong Huang, Zhi-Qiang Mi, Dong Yu,
Hong-Wu Yao, Yong Feng, Zhi-Ping Xia, Xue-Xing Zheng,
Song-Tao Yang, Bing Lu, Jia-Fu Jiang, Brima Kargbo,
Fu-Chu He, George F. Gao, Wu-Chun Cao
& The China Mobile Laboratory Testing Team in Sierra Leone

This Letter should have contained an associated Creative Commons statement in the Author Information section. In addition, the Fig. 3c legend should have stated that the bar chart was adapted, with permission, from Ebola response roadmap - Situation report, Figure 3; http://www.who.int/csr/disease/ebola/situation-reports/en (accessed 1 April 2015)[1]. These issues have now both been corrected in the online versions of the paper.

1. World Health Organization. Ebola response roadmap - Situation report. http://www.who.int/csr/disease/ebola/situation-reports/en (accessed 1 April 2015).

CORRECTIONS & AMENDMENTS

Erratum: Evidence for human transmission of amyloid-β pathology and cerebral amyloid angiopathy

Zane Jaunmuktane, Simon Mead, Matthew Ellis,
Jonathan D. F. Wadsworth, Andrew J. Nicoll, Joanna Kenny,
Francesca Launchbury, Jacqueline Linehan,
Angela Richard-Loendt, A. Sarah Walker, Peter Rudge,
John Collinge & Sebastian Brandner

In this Letter, an administrative error led to the publication of an incorrect version of the Competing Financial Interests (CFI) statement. Although the published CFI statement did reference the authors' affiliation with D-Gen, it did not contain all of the information provided by the authors about the interests of the company. The CFI statement for this paper as originally published was "J.C. is a Director and J.C. and J.D.F.W. are shareholders of D-Gen Limited, which supplies antibody ICSM35." The updated CFI statement is "J.C. is a Director and J.C. and J.D.F.W. are shareholders of D-Gen Limited, an academic spin-out company working in the field of prion disease diagnosis, decontamination and therapeutics. D-Gen supplied antibody ICSM35."

CORRECTIONS & AMENDMENTS

Corrigendum: Lanosterol reverses protein aggregation in cataracts

Ling Zhao, Xiang-Jun Chen, Jie Zhu, Yi-Bo Xi, Xu Yang, Li-Dan Hu, Hong Ouyang, Sherrina H. Patel, Xin Jin, Danni Lin, Frances Wu, Ken Flagg, Huimin Cai, Gen Li, Guiqun Cao, Ying Lin, Daniel Chen, Cindy Wen, Christopher Chung, Yandong Wang, Austin Qiu, Emily Yeh, Wenqiu Wang, Xun Hu, Seanna Grob, Ruben Abagyan, Zhiguang Su, Harry Christianto Tjondro, Xi-Juan Zhao, Hongrong Luo, Rui Hou, J. Jefferson P. Perry, Weiwei Gao, Igor Kozak, David Granet, Yingrui Li, Xiaodong Sun, Jun Wang, Liangfang Zhang, Yizhi Liu, Yong-Bin Yan & Kang Zhang

In this Letter, author Yong-Bin Yan was incorrectly associated with affiliation number 5 (Department of Ophthalmology, Xijing Hospital) instead of affiliation number 4 (State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China). Also, an additional affiliation has been added to author Kang Zhang (number 15; Institute of Molecular Medicine, Peking University, Beijing 100871, China), and affiliation number 3 has changed from 'Department of Ophthalmology and Biomaterials and Tissue Engineering Center' to 'Shiley Eye Institute and Biomaterials and Tissue Engineering Center'. These have all been corrected in the online versions of the paper.

CORRECTIONS & AMENDMENTS

Corrigendum: Selective killing of cancer cells by a small molecule targeting the stress response to ROS

Lakshmi Raj, Takao Ide, Aditi U. Gurkar, Michael Foley, Monica Schenone, Xiaoyu Li, Nicola J. Tolliday, Todd R. Golub, Steven A. Carr, Alykhan F. Shamji, Andrew M. Stern, Anna Mandinova, Stuart L. Schreiber & Sam W. Lee

In this Letter, we presented findings from experiments using the EJ bladder xenograft cancer model, in which some tumours on some of the animals exceeded the maximum size (15 mm in any dimension) permitted by the Institutional Animal Care and Use Committee (IACUC) at Massachusetts General Hospital (MGH). Therefore, we withdraw the data presented in Supplementary Fig. 9b, as well as in Fig. 2 from the first Corrigendum. Although other tumours were found to exceed the permitted maximum, owing to the degree of tumour size and the circumstances of the experimental procedures, this was considered acceptable as detailed below.

The tumours in eight mice in Fig. 2b, d (same mice in Fig. 1a, b in the Corrigendum) also exceeded the tumour size approved in the IACUC protocol of the principal investigator (PI), and were euthanized 48–72 h after the tumour burden was identified. Although the tumour size limit permitted by the protocol had been reached in the above-mentioned animals, the mice exhibited no other clinical signs associated with humane endpoints due to pain or distress. Given the lack of clinical signs observed and the timing of the clinical presentation, the PI was permitted to maintain the animals for this short time period to support proper tumour collection and fixation. The animals were regularly monitored during this time period. We also noted that we made an error in the first Corrigendum when calculating the volumes for the last time point of Fig. 1b. All measurements are now presented as Supplementary Data.

The *in vivo* experiments described in Supplementary Fig. 9 were performed between 2007 and 2008. At this time, the IACUC-approved protocol of the PI required daily monitoring of the tumour-bearing animals for clinical signs of distress, and did not require daily measurements of the tumour size. Measurements of tumour sizes were performed at the indicated time points. Consistent with the IACUC-approved guidelines, all animals were euthanized as soon as measurements indicated that the tumours reached the size limit approved in the protocol (15 mm). These mice did not show clinical signs of distress, and thus it only became apparent that the tumours had already exceeded 15 mm when they were measured. Since the completion of this study, the IACUC-approved protocol of the PI was revised, and daily measurements of tumour sizes, in addition to daily observation, are now required. For the xenograft tumour models, measurements were performed on the entire tumour lesion including cases when tumours appeared as aggregates of single nodules (melanoma model in Supplementary Fig. 9e). All measurements are now presented as Supplementary Data.

The MGH IACUC approved the animal models and the general procedures for drug testing used in this work; however, some compounds and cell lines used for the experiments in Fig. 2 and Supplementary Fig. 9 had not received prospective IACUC approval owing to an administrative oversight. All of the materials included in the original Letter and the earlier Corrigendum have since been approved by the IACUC for subsequent experiments. We would like to clarify that the detection of total p53 in Fig. 4c was performed after stripping the membrane following detection of phosphorylated p53 (Ser15 p-p53). In addition, the original paper incorrectly stated that the error bars in Figs 1c, 3a–d, 4a, b, and Supplementary Figs 2b, 3a, 4a, 6, 15c, 18, 20, 21, 26b and 29b were calculated based on three independent experiments. These graphs and the associated error bars represent the results of technical triplicates from one experiment. The error bars in the graphs throughout the main paper and Supplementary Information represent the standard deviation of the mean. The IACUC has reviewed all the data now presented, and confirms that the statements provided in this Corrigendum are accurate. Corrective measures have since been taken to avoid any irregularities happening again. Although the scientific conclusions of the original paper stand, we would like to apologize for the numerous inaccuracies in reporting our data, and for the breach of animal welfare guidelines in some of the original data.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

PRIME TIME

Parental control.

BY JENNIFER CAMPBELL-HICKS

Aurelia's Dad stood on the basement stairs, his electrically charged hair bristled like porcupine quills. Behind him stood three more Dads who were exact copies of the first.

"Aurelia, thank goodness," said the Dads. "I'm still at —"

"The movie with Mom? Yep."

"Good. I used my time machine to come back and warn you. Tonight, before Mom and I get home, you will vanish without a trace." The Dad in front grabbed her shoulders. "Fate is not set. I'm here to change it. And stop biting your nails."

She snatched her fingers from her mouth. It was a coping mechanism. Senior year, harder classes, applying to colleges. There were no strategies for coping with this.

"You're 11A," she said. "11B, 11C, 11D."

They looked confused. "What?"

"Just follow me."

She led the four Dads to the living room. Six more copies sat on the leather couch, the love seat and the recliner. One leaned against the wall. They looked anxious. One Dad — 3A, she thought — clicked the remote at the television.

"What's this?" the 11s asked.

"What do you think? Your machine is broken. It's spitting you out, over and over. You're coming out in groups so you always add up to a prime number. We had seven. Now it's eleven." She was proud that she had seen the pattern and gave credit to four years on the school maths team.

"Primes? Why?"

"How should I know? It's *your* stupid machine. And before you ask, because you've already asked, there are 10 of you here because 1A is in the basement trying to shut down the machine." Her gaze slid over them. "You should have listened to Mom. She said it was a bad idea to mess with the space-time continuum."

Knocking shook the basement door.

"What was that?" a Dad asked.

"You," Aurelia said and gnawed at her thumbnail. "Have a seat. If there's one left."

Back in the kitchen, she opened the basement door. This time, as she expected, two Dads stood on the steps.

"Aurelia, thank goodness —"

"I know. I vanished. You're here to save me." She pointed. "13A and 13B. Come with me."

If this continued, dozens of Dads might be all over the house when her parents got home. She ran the numbers: 17, 19, 23, 29, 31, 37, 41, 43 …

"What's this?" the 13s asked in the living room.

"The others will explain. I need a volunteer."

"Me," they all said.

She pointed. "Who are you?"

"11B."

"Come on."

In the basement, the time machine hummed and clicked. It was an 8-foot cube of metal, gears, wires and pulsating light with an open hatch. Soon, if nothing changed, four more Dads would emerge from the hatch. The timing wasn't regular. They could appear 10 seconds from now or 10 minutes. Dad 1A peered out from behind the machine, grease on his cheeks.

"There's something wrong, obviously," he said. "I can't shut off the power."

"It'll keep spitting you out?" Aurelia asked. He nodded.

"For how long? Eternity?"

"Or until the space-time continuum overloads. Can it do that? I don't know." He shook his head. "I should have listened to your mother. I'm sorry."

He looked so sad that Aurelia hugged him.

"I wish we had more time," 11B said.

That gave Aurelia an idea. "We do have time. I can jump forward and tell you not to jump back." Then she thought of something else. "Wait. That must be how I vanished in the first place, which was why you came back to save me."

11B stepped forward. "I'll go."

"We can't meet ourself like that," said 1A. "Paradox risk."

Aurelia chewed on two fingernails at once. Was this how the apocalypse started? The world became overrun by crazy-haired clones of her Dad? Then she remembered what he had said on the basement steps. Fate was not set. They could change it.

"You can't shut off the machine," she said. "Can you send it to another time?"

"I don't know," Dad 1A said, sounding dubious, but Dad 11B grew excited. "We can program it so it's always five minutes from the present! If it's always five minutes in our future, it will never overrun our now with clones."

"And if I don't go to warn you," Aurelia added, "I won't disappear, which means you won't come here at all."

It was perfect. They could solve the problem before it started, but Dad 1A's shoulders slumped.

"If we send the machine to the future, I'll lose it. All those years of work will be for nothing."

"I'm sorry," Aurelia said. "It doesn't work right, anyway. The alternative is Dad clones coming out in primes into infinity. The Earth can't handle that."

He nodded. The two Dads got to work with the machine's controls while Aurelia watched with her hands stuck firmly under her armpits so she wouldn't chew on her nails. Then they said: "Ready to rock and roll. Be good, kiddo."

They pressed a button. The machine hummed louder, glowed brighter. Aurelia squeezed her eyes shut. When she opened them, the machine was gone, leaving a square of clean concrete floor amid the dust. The Dads were gone, too.

She sprinted upstairs. The living room was empty, thank goodness. Then a key turned in the front door. Aurelia's fingers rose to her mouth, but she stopped and put her hands at her sides. She could handle this. The door swung open. Dad walked in with Mom behind him, both in nice date-night clothes.

"Aurelia," Dad said. "It's a school night. Why aren't you in bed?"

She took a deep breath. "Dad, do you have time to talk?" ∎

Jennifer Campbell-Hicks *is a writer, journalist, wife, mother and lifelong fan of science fiction and fantasy. Her fiction has appeared in* Daily Science Fiction, Flash Fiction Online *and* Intergalactic Medicine Show. *She blogs at jennifercampbellhicks. blogspot.com.*

ILLUSTRATION BY JACEY
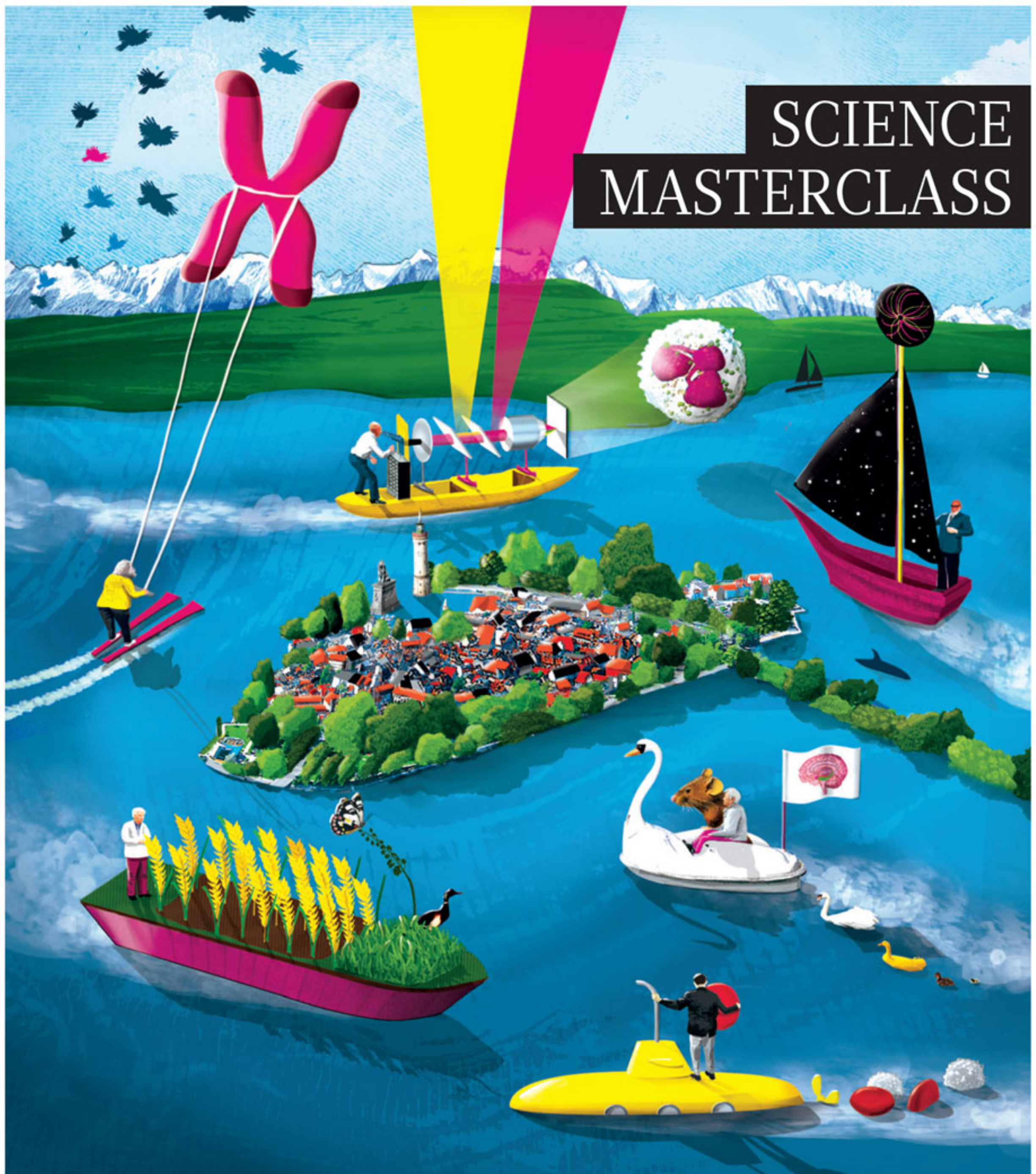
SCIENCE MASTERCLASS

Ideas in motion

natureOUTLOOK

SCIENCE MASTERCLASS

22 October 2015 / Vol 526 / Issue No 7574

Cover art: Sam Falconer

Lindau is a tiny island, covering two-thirds of a square kilometre. For most of the year, its medieval streets are filled with tourists. For one week in the summer, however, the Bavarian town becomes home to an exchange of ideas that circulate beyond its limits and, in the future, could stretch even further. This is because Lindau is home to the Nobel laureate meetings where past prize winners meet young researchers hoping to pick up wherever their illustrious mentors leave off.

The annual event began in 1951, two years before James Watson and Francis Crick published the structure of DNA. In 2015, it welcomed Elizabeth Blackburn (see page S56), who studies the genetic material at the ends of chromosomes, and Richard Roberts (S58), who found that genes often contain non-protein-coding portions that facilitate the creation of many different proteins from a single gene.

Similarly, by the first Lindau meeting, physicists had not unified electromagnetism and nuclear weak forces. One of this year's guests included François Englert (S61), who imagined the Higgs boson. Laureates Bruce Beutler (S59) and Susumu Tonegawa (S55) — specialists in cellular immunity and emotional memories, respectively — are also furthering fields that barely existed in 1951. This Outlook illuminates all of these Nobel prize winners' discoveries. An in-depth look at super-resolution microscopy — the subject that opened this year's Lindau meeting — describes the impact that the technology is having on molecular and cell biology (S50). This feature and the Q&A with Blackburn are accompanied by an animation; you can also hear the Q&As with Tonegawa and Beutler as a podcast (all available at www.nature.com/outlook/masterclass2015).

We are pleased to acknowledge financial support from Mars, Incorporated in producing this Outlook. As always, *Nature* has sole responsibility for all editorial content.

Anna Petherick
Contributing editor

CONTENTS

$$d = \frac{\lambda}{2n\sin\alpha}$$

The equation describing the diffraction limit is engraved on a monument to optical physicist Ernst Abbe in Jena, Germany.

CELL IMAGING

Beyond the limits

Powerful super-resolution microscopes that allow researchers to explore the world at the nanoscale are set to transform our understanding of the cell.

KATHERINE BOURZAC

Dyche Mullins dons a pair of Oculus Rift 3D goggles and is transported from a cubicle at the University of California, San Francisco, into a virtual world. Through the virtual-reality headset, Mullins watches an immune cell that appears to be the size of a child. The cell crawls through a maze of collagen fibres. Mullins can move around the image, 'pushing' in so that the neutrophil is crawling directly above him, or turn the image around to watch the cell's motions from different angles.

This is not a simulation. The video was made using a cutting-edge microscope invented by Eric Betzig, an engineer at the Howard Hughes Medical Institute's Janelia Research Campus in Ashburn, Virginia. Foundational work in cell migration is based on studies done in the 1950s and 1960s in 2D, at low resolution and on glass slides. Seeing cells more clearly, says Mullins, who is a molecular biologist, is helping to overturn received wisdom about the fundamentals of how cells move, whether they are immune cells moving through tissue or amoebae in a pond. With new imaging tools, he can watch their movement in vivid detail.

These tools are redefining microscopy's frontier, and their potential impact on biology is huge. Until the development of super-resolution microscopes, which began in 2000, life's underlying molecular world appeared as a blur, only hinted at by relatively low-resolution light

microscopy or captured in crisp, but static single-frame electron micrographs. "Now we have the proof that these barriers can be overcome," says Stefan Hell, director at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. "We're just at the beginning of what's possible." Hell, a newly minted Nobel laureate, gave a lecture on the subject this year at the 65th Lindau Nobel Laureate Meeting in Germany, rousing excitement among his fellow laureates and the young scientists in attendance.

Hell was jointly awarded the Nobel Prize in Chemistry in 2014 with Betzig and William Moerner for their development of

DANIEL MIETCHEN

super-resolved fluorescence microscopy. Before their work, the nanoscale details of life in motion went unwitnessed. That has been a big blind spot. Life is dynamic: cells not only crawl around, like Mullins' neutrophil, they divide, shuttle chemicals from one internal structure, or organelle, to another, and so much more. Aside from cell migration, researchers are using super-resolution microscopes to look at what happens when a group of viruses attacks a cell in real time, and they are turning these instruments on the chemical connections between neurons in the brain. "This technology will be transformative," says Tomas Kirchhausen, a cell biologist at Harvard Medical School in Boston, Massachusetts, and another early adopter of these tools. "I'm ready to put my other instruments on eBay."

A SHARPER IMAGE

Hell was the first to break the existing limits on light microscopes. But not too long ago, he told the Lindau audience, no one took him seriously (a common remark made by Nobel prize winners). Ideas about the limits of imaging were not only set in scientists' minds, but literally set in stone. On the lecture screen, Hell showed a photograph of a monument to one of his scientific forebears, Ernst Abbe, in Jena, Germany — a stone engraved with the equation describing the diffraction limit.

Abbe rose from humble beginnings to become a brilliant optical physicist at German microscope manufacturer Zeiss. He was the first to understand how images are made in light microscopes, explained Hell, and he was able to make the best instruments. Abbe figured out that when a beam of light is focused on a spot whose size is smaller than about half its wavelength, it starts to interfere with itself, and the image becomes blurred. If two objects are closer together than that distance — a couple of hundred nanometres — they cannot be resolved under a conventional light microscope no matter how perfect the lens. This is Abbe's diffraction limit.

It is in large part thanks to Abbe's work that light microscopes have been powerful research tools. "I have huge respect for Abbe," says Hell, who named his own microscopy company, based in Göttingen, Germany, Abberior. Abbe was not wrong, he was a product of his time. In 1873, when he published his work on the diffraction limit, the molecule was an unproven idea and making better microscopes was all about developments in optical science.

Hell realized that chemistry could take things further. Below the diffraction limit is a rich microbiological world that includes viruses, proteins and the fine details of organelles such as the energy-generating mitochondria. Hell was determined to bring this world to light, and became obsessed with breaking the diffraction limit. He focused on a technique called fluorescence microscopy, a common tool used by biologists. The method
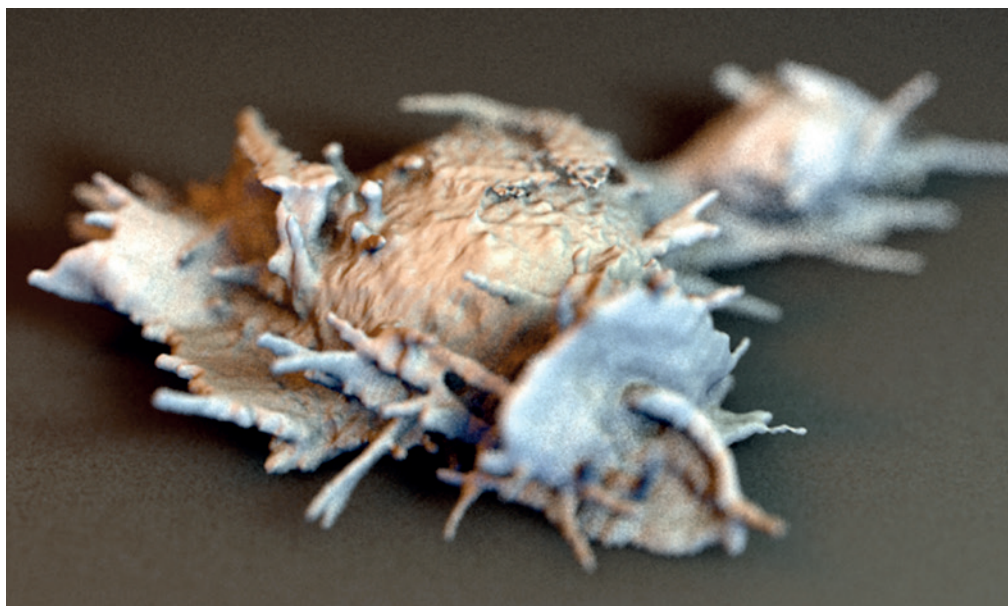
A live cell captured by a type of microscopy developed by Nobel prizewinner Eric Betzig.

is a form of light microscopy, augmented by fluorescent chemical labels. These bright red, green or other coloured tags attach to the specific molecules that a biologist is interested in and act as beacons to help locate particular types of cells or structures within cells.

In the 1990s, Hell read an article about a quantum mechanical phenomenon called stimulated emission — a way to control the fluorescence of these labels — and was "electrified". He realized that imaging could be about more than just great lenses — he could use chemistry to get beyond the diffraction limit.

Hell designed a system that used two beams of light, one to stimulate fluorescent molecules, and another that immediately turned most of them off in such a way that only those at the centre of the light beam continued to shine[1]. The technique, which he called stimulated emission depletion (STED) microscopy, does not break the diffraction limit but, as Hell puts it, plays a game to get around it. The resulting cylindrical beam — shaped like the outline of a donut with the inner circle filled in — is then scanned over the sample to get the full picture (see 'Under the microscope'). Hell's first super-resolution microscope in 2000 only exceeded the diffraction limit by around a factor of two, but it brought the barrier down.

Hell is a self-described dreamer, and felt shut out of the mainstream of science before his Nobel-winning discovery. Whereas Betzig cultivates the aura of a cranky outsider, and in his talk at Lindau this year, he made it clear that he has no patience for the strictures of academia. Betzig calls himself a "tool-builder" — an engineer who wants to build the best microscopes possible.

Betzig was also long-obsessed with surpassing Abbe's diffraction limit. Working on near-field microscopy at renowned research centre Bell Labs in Murray Hill, New Jersey,

in the early 1990s, he made advances on a technology called near-field imaging. But he grew frustrated with research on what he did not consider to be a very useful tool. Near-field microscopy technically beats the diffraction limit, but it is not very practical because it uses a sharp imaging tip that must be extremely close to the sample. Betzig left research for several years and worked at his father's company, designing machine-tooling equipment. After inventing a design that the company could not sell, he had what he calls his second midlife crisis. He kept up-to-date with the scientific literature, and reading about green fluorescent proteins brought him back to science — into the same sphere in which Moerner was making waves — and so he returned to the problem of the diffraction limit.

> *"The technology will be transformative. I'm ready to put my other instruments on eBay."*

Like Hell's solution, Betzig's relies on chemistry, in this case a phenomenon called photoswitching that Moerner was studying[2]. Working at IBM in San Jose, California, Moerner was the first to measure the absorption spectrum of a single molecule[3]. Then, in 1997, he went on to show that certain fluorescent molecules could be turned on and off with a beam of light. The underlying physics are different from STED, but the implication is the same: being able to turn fluorescent molecules on and off as though they were light bulbs.

Hell's microscopes raster a narrow beam over a sample to fill in the image zone by zone, whereas Betzig's work combines several faint snaps of the whole imaging zone to make a complete picture. In the latter case, before each image capture, the microscope illuminates
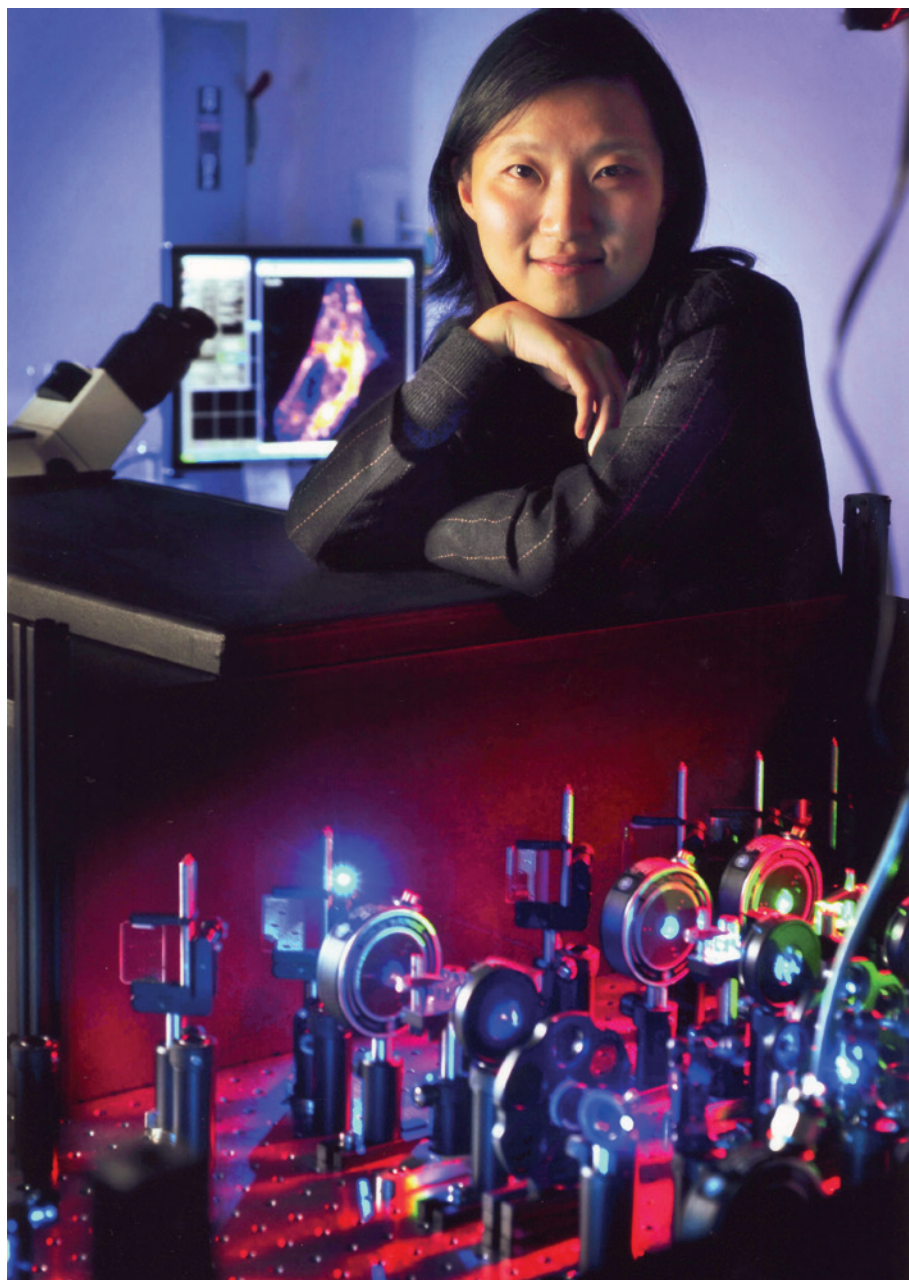
Xiaowei Zhuang developed the super-resolution microscopy method STORM.

the sample with a weak beam of light that only turns on a small fraction of the fluorescent molecules in the sample. This is repeated again and again until all of the molecules have been located. Even if two molecules are closer together than the diffraction limit, they will appear in different sub-images, so they can all be seen in composite.

CELLS IN FOCUS

Biologists trying out the new technologies are amazed by what is possible. In the seventeenth century, before he examined a slice of cork that he had put under his microscope, Robert Hooke did not know that cells existed — no one did. But there they were: "… indeed the first microscopical pores I ever saw, and perhaps, that were ever seen, for I had not met with any

Writer or Person, that had made any mention of them before this," he wrote in 1665. Stories from early users of super-resolution imaging are reminiscent of Hooke. "Super-resolution is not just about getting sharper pictures," says Xiaowei Zhuang, a biophysicist at Harvard University in Cambridge, Massachusetts, who developed another super-resolution method called Stochastic Optical Reconstruction Microscopy (STORM), which she and her colleagues published in 2006 (ref. 4). "You can discover things that have never been seen before."

Using STORM to look at the structural protein actin in neurons in 2012, she saw something surprising and new. Zhuang is interested in synapses, the electrochemical connections between neurons that are the basic units of brain circuits. She had asked her postdoctoral

researchers to image the structure of synapses by mapping out actin at high resolution. But the most noticeable feature of the pictures turned out to lie next to the synapses: in the images, the axons — the slender projections that transmit neurons' outgoing signals — looked odd.

All along the axons, the team saw regularly spaced rings of actin. "It's a remarkable periodic structure — every 180 nanometres, an actin ring," says Zhuang. "It's beautifully laid out, it almost looks like the structure of a crystal." The features of this pattern are under the diffraction limit, and so the rings had been previously invisible. Under a conventional fluorescent microscope you can see a smear indicating the presence of actin, but the periodic rings, so distinct under Zhuang's STORM microscope, are a blur. Nor had anyone noticed the rings under an electron microscope because other filamentous structures in the cell obscure them, she says.

Zhuang has found that the actin rings are present in all axons, but only in patches on a small fraction of dendrites, the neural projections that receive signals. The 'axonal skeleton' helps axons to maintain their mechanical stability and conduct impulses. This understanding is supported by other studies[5], which show that worms genetically engineered to lack the protein that connect the rings have fragile axons, impaired movement and a reduced response to mechanical stimuli.

Harvard's Kirchhausen has had similarly eye-opening moments using a microscope built with Betzig, which uses thin sheets of light to illuminate one planar slice of a sample at a time and image it using Betzig's super-resolution method. It is the most recent example of Betzig's work. Kirchhausen has spent decades studying the formation of membrane bubbles called vesicles that are just 30–80 nanometres in diameter. Vesicles carry chemicals around the cell and also help the cell engulf, transport and break down unwanted molecules and microbes. In 2004, Kirchhausen completed the first image of the structure of the protein complex that pinches off vesicles from a membrane[6]. He used electron microscopy to generate images with near-atomic resolution,

> *"Super-resolution is not just about getting sharper pictures. You can discover things that have never been seen before."*

going down to 0.8 nanometres. But although he was making progress in determining the structure of vesicles, he says, "I realized I should be paying attention to the dynamics."

Using one of Betzig's machines, he was finally able to watch the process. Over a period of 30–90 seconds, hundreds of molecules assemble into a kind of geodesic dome that pulls out a bubble of membrane. Then they all fall away to leave the vesicle. Kirchhausen can now watch

UNDER THE MICROSCOPE

Our understanding of cells such as neurons has benefited from advances in imaging. By **Katherine Bourzac**.

LIGHT MICROSCOPY SIXTEENTH CENTURY

The first light microscopes used relatively simple arrangements of lenses to magnify tissue; later versions used external illumination and stains to make features even clearer. The original inventor is disputed, but the first microscopes are thought to have emerged from work on telescopes.

ELECTRON MICROSCOPY 1933

Physicist Ernst Ruska created the first microscope to form an image by bouncing beams of electrons off a sample. He shared the 1986 Nobel Prize in Physics for this work. Electron microscopes can resolve atomic-scale detail — as static, structural images — because the wavelength of an electron is much shorter than that of light.

FLUORESCENCE MICROSCOPY 1990s

This technique allows researchers to locate parts of the cell using chemistry. Molecules of interest can be labelled by attaching fluorescent tags to proteins, and gene expression can be monitored by inserting a fluorescent protein gene into the genome (a method recognized by the Nobel Prize in Chemistry in 2008).

SUPER RESOLUTION 2000 onwards

The latest microscopes take resolution to the nanoscale. The subject of the 2014 Nobel Prize in Chemistry, super-resolution microscopy zooms in on individual molecules and subcellular structures by computationally combining many fluorescence images. Stimulated emission depletion, or STED, microscopy is an example of this technique.

Axon terminal

Myelin

Schwann cell

Nucleus

Cell body
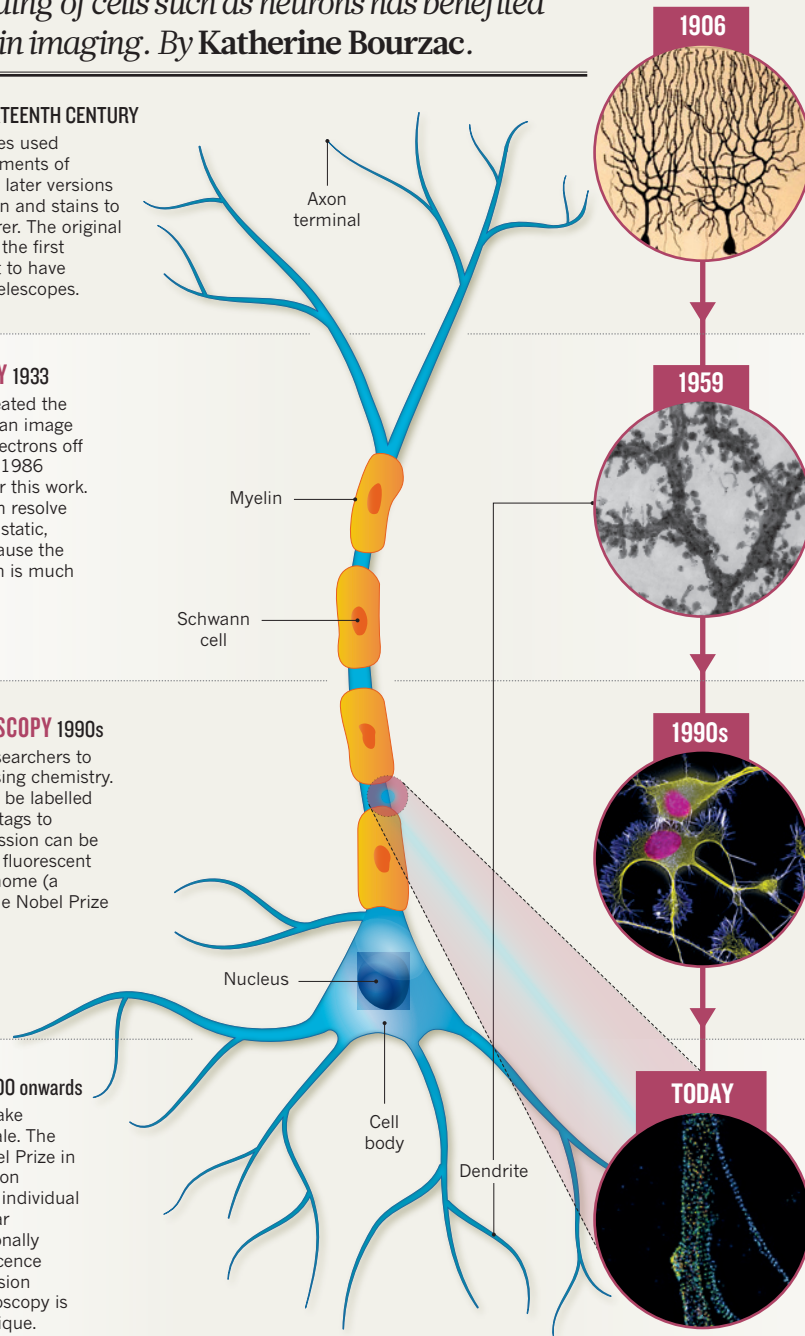
Dendrite

1906

Under a light microscope, neuroscientist Santiago Ramón y Cajal saw the delicate branching of dendrites. He shared the Nobel Prize in Physiology or Medicine in 1906 with Camillo Golgi for their work on the structure of the nervous system.

1959

Electron microscopes revealed the branching or 'spines' on the dendrites of neurons for the first time. In 1959, Edward George Gray published images of dendritic spines and established that the connections between neurons (synapses) occur at these sites.

1990s

Fluorescent imaging advanced our understanding of the cell by revealing how the protein actin helps in cell growth and migration. Labelling actin with fluorescent dyes, researchers saw that it propels the leading edge of growth in axons, which carry neurons' outgoing electrochemical signals.

TODAY

In 2013, biophysicist Xiaowei Zhuang used super-resolution microscopy to show that axons have rings of actin along their length, like tiny, regularly-spaced belts. These belts are believed to provide structural support.

LIFE IN FOCUS

Under a conventional optical microscope, the resolution is limited to a few hundred nanometres, and the details of a mitochondrion are blurry. STED microscopy uses an exciting beam to image a sample and a quenching beam to cancel out part of the beam, narrowing it and increasing the resolution.

Optical microscope

Exciting laser beam

STED microscopy

Quenching laser beam

Exciting laser beam

1

2

3
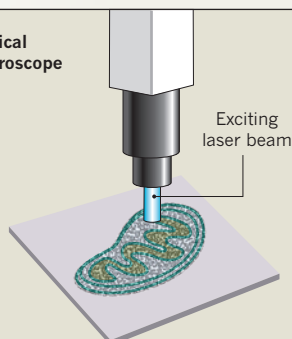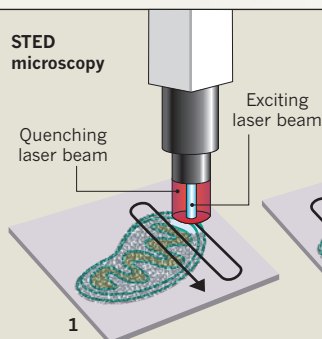
At a resolution of tens of nanometres, the features of the mitochondrion become visible, revealing the folds of the organelle's internal structure.
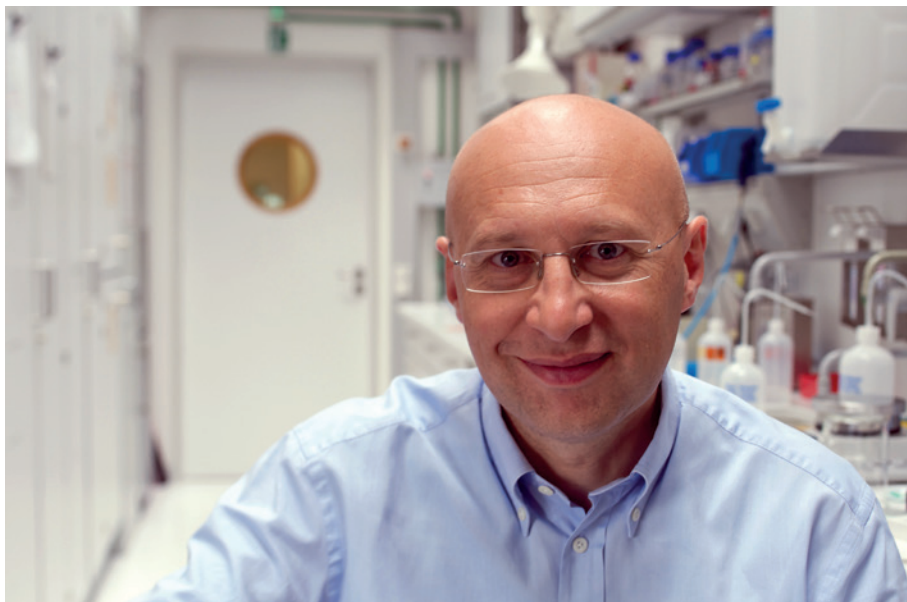
Stefan Hell was jointly awarded the Nobel prize for his work on super-resolution microscopy.

this happen thousands of times a minute, throughout a cell.

As soon as the super-resolution microscope was up and running in 2014, Kirchhausen started using it for various other projects requiring super-resolution in 3D. "We sprinkle viruses on a cell and watch how they're taken in," he says. "Now we can track every virus as it enters the cell and follow its fate." Which viruses succeed? Which don't? And what happens as a virus enters a cell? Those are questions Kirchhausen can now ask.

CHALLENGES AHEAD

Using super-resolution microscopes is not like working an old-fashioned light microscope. "There are no binoculars. Everything is done in the computer," says Kirchhausen. "You can see a crude version of what is coming off the microscope, but it takes some time afterwards to render it into something useful," says Mullins. These methods are based on imaging huge numbers of single molecules at a time. That is something that neither the human eye nor brain can make sense of unless powerful software has processed it first.

Most image-processing software has been designed to work with 2D, low-resolution images, explains Mullins. For his work, he collaborated with specialists in processing 3D images (previously for computer models of individual molecules such as proteins) at the University of California. The video of the neutrophil crawling through a couple of micrometres of tissue was acquired in just 3 minutes in a single experiment that yielded 25 gigabytes of data. A typical high-definition Hollywood film is about one-tenth of that size. Data sharing and storage are becoming a problem.

Hardware is another barrier to using super-resolution microscopy. Moerner is using super-resolution microscopes based on both

Betzig and Hell's designs in his lab. His set-ups are complex. A square metre or so of about 30 precisely placed optical components direct the imaging light. The microscopes have to be housed on a vibration-isolating table that is common in physics departments, but not in wet labs.

Such vulnerability is not ideal. Companies including Zeiss and Hell's start-up are designing commercial models that put more compact, fixed versions of these set-ups inside a case. So far the models on the market have not been very good, says Betzig, but that will change.

This move towards commercialization does not signal that the fundamental research is slowing down. Hell's dream now, he says, is to map every protein in living tissue in real time at a resolution of 1 nanometre, without causing the tissue any damage. Zhuang's dream is similar. One goal in her lab is to follow the waxing and waning of gene expression at the genome scale in individual cells — that would mean imaging the full orchestra of messenger RNA in cells at the same time. This April, she published results from her first attempt, which used a combination of single-molecule imaging and computational savvy to simultaneously detect about 1,000 different kinds of RNA molecules in one cell[7]. Because she is interested in the brain, Zhuang particularly wants to watch how neurons develop, to understand how their different behaviour emerges from gene expression patterns.

Another of Zhuang's goals is to track communication between neurons at their synapses, to watch these cell-to-cell chemical communications just as electrophysiologists

"Until I watched this, I didn't understand the psychology of the cell."

can measure a single neuron's electrical signals. One of her projects is to follow computation in individual neurons. Currently, no one fully understands how a neuron integrates incoming signals and then generates an output. Once that question can be answered at the level of a single cell, observations can be turned towards figuring out the processes of computation in small circuits, and eventually in the whole brain.
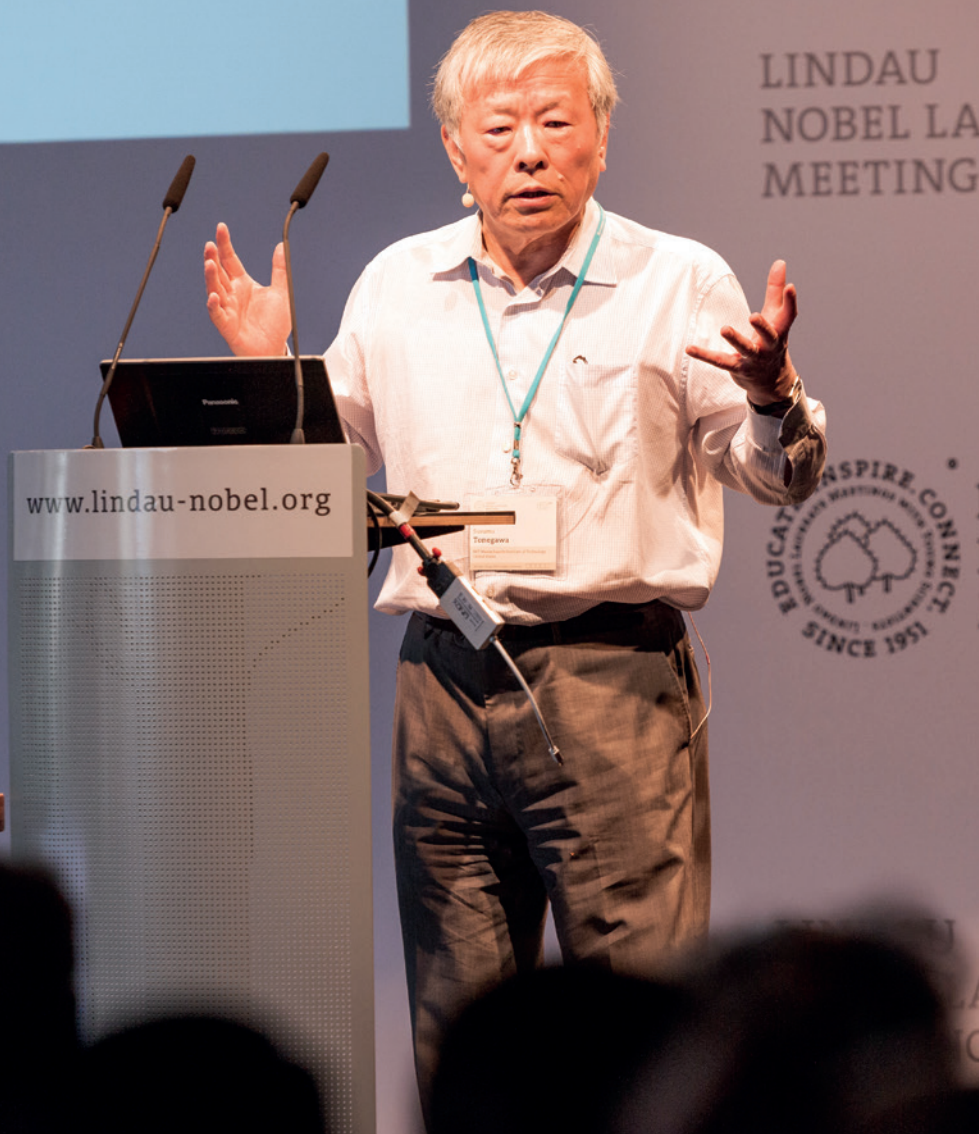
Imaging individual cells with these methods is now possible, but it is more difficult to take speedy, single-molecule resolution videos at the tissue level, for example in the brain. Today's systems are held back by the brightness and switching response of existing imaging dyes, and the limited number of colours that can be used at once.

To help them advance the field, Zhuang, Hell, and other imaging researchers are looking to chemistry. "The next frontier is going to be new dye colours," says Luke Lavis, a chemist at the Janelia Research Campus who develops fluorescent imaging tags. Creating dyes that attach to biomolecules in ways that disrupt their function as little as possible — and making a whole rainbow of dyes with colours distinct enough to be separated during image processing — will make it possible to view more moving molecules simultaneously. And making dyes brighter, or designing them so that they will not fluoresce until they attach to their target, might stop cells from becoming badly damaged by repeated blasts of high-energy-stimulating light. Lavis is working on it. "When I started out, people would be using synthesis methods reported in 1888, but chemistry has evolved so much since then. Now we can really explore the structures of these dyes," he says.

Watching Mullins explore the subcellular world in his 3D goggles, it is clear that the details unveiled by today's microscopes would astound researchers such as Abbe and Hooke. Nonetheless, super-resolution imaging is not all about seeing and quantifying cells as collections of molecules. "Until I watched this, I didn't understand the psychology of the cell," says Mullins, only half joking. In this footage, the cell is not an abstract object. It responds to mechanical forces and it pushes back. The cell has a palpable physical presence. And that, compared to all means of watching it before, somehow makes it seem more alive. ∎

Katherine Bourzac *is a science journalist based in San Francisco.*

1. Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. *Proc. Natl Acad. Sci. USA* **97,** 8206–8210 (2000).
2. Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. *Nature* **388,** 355–358 (1997).
3. Moerner, W. E. & Kador, L. *Phys. Rev. Lett.* **62,** 2535–2538 (1989).
4. Rust, M. J., Bates, M. & Zhuang, X. *Nature Methods* **3,** 793–796 (2006).
5. Zhong, G. *eLife* **3,** e04581 (2014).
6. Fotin, A., *et al. Nature* **432,** 573–579 (2004).
7. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. *Science* **348,** 6233 (2015).

Q&A Susumu Tonegawa

Memory man

Susumu Tonegawa unlocked the genetic secrets behind antibodies' diverse structures, which earned him the Nobel Prize in Physiology or Medicine in 1987. Having since moved fields, he tells Keikantse Matlhagela about his latest work on the neuroscience of happy and sad memories.

You started as a chemist, then you moved into molecular biology and now you are a neuroscientist. Why change fields?
Strangely, the only people to ask me about this are journalists — my students never ask. I see myself as a scientist who is interested in what's going on inside of us. It doesn't matter whether it is chemistry or immunology or neuroscience, I just do research on what I find interesting. The switch from chemistry to immunology did not seem like a big shift when I was young, but immunology to neuroscience was. After about 15 years spent researching immunology I wanted to explore an area of science where there are still big, unresolved questions. The brain is probably the most mysterious subject there is.

Do you keep up to date with the field in which you won your Nobel prize?
I am sorry to say that I haven't been paying a lot of attention to immunology in recent years because I am preoccupied with my work on memory. I have friends, of course, from that time — very close friends. But my friends are not young. Even though they are experts, they are also retired. We tend not to talk about immunology a whole lot.

Is it helpful in neuroscience research to have a multidisciplinary background?
The brain is hugely complicated, and because it is so complicated it requires multidisciplinary research. You need mathematics to model how brain networks function. You need chemistry, molecular biology and behavioural studies of animals to answer other neuroscience questions. Neuroscience is totally multidisciplinary.

What have you learnt about the brain circuits involved in positive and negative emotions?
Imagine that a week ago you were on a vacation — you went to a Caribbean island and had a great time — and you remember the detail of what happened during your vacation. Those memories would be examples of 'episodic memory'. Sometimes episodes come with no emotional content, but often they come with a positive or negative slant — in other words, they were either pleasurable or unpleasant experiences.

My lab has been studying the part of the brain called the hippocampus to investigate its role in the formation of episodic memory, and how that varies with positive and negative emotional content. Our results indicate that there is a kind of competition between brain circuits to be able to assign a positive or a negative value to the memories. We have taken advantage of this understanding in our experiments on mice and have shown how depression can be reversed or repressed.

How do you tell if a mouse is depressed?
It is similar to the way in which you tell if a human is depressed. There are at least two symptoms. When a depressed patient encounters something difficult to resolve or improve, they give up more easily than people without depression. Another symptom is called anhedonia, which is an inability to enjoy normally pleasurable experiences. Therefore, depressed people don't seek out normally pleasurable experiences. The same goes for mice with depression.

Have you identified a target protein or group of proteins involved in mouse depression?
No. But we have found specific target cells that hold pleasurable episodic memories. They are deep inside the brain. There is an area of the brain that in male mice, for example, holds information about a specific, playful encounter with female mice. We have developed technology to identify these cells and we have genetically manipulated mice so that the cells express a light-sensitive protein. If you shine blue light onto them the cells become activated, meaning that the mouse recalls the positive experience. Going back to the depression model: if a mouse is depressed, activating these previously identified pleasure-memory cells can cure its depression. So when

⟳ NATURE.COM
To listen to the interview visit:
go.nature.com/wcwqdj

you put the mouse in a difficult situation that it would have previously given up on, it now makes an effort to solve the problem.

Have you tested this in higher animals?
Not yet. Unfortunately, the technology that we use for mice is not directly applicable to humans. But I know researchers are working on how to replicate in humans something similar to what we did in mice and eventually we can expect that this will become possible. Hopefully, one day, our findings will lead to a new kind of therapy for depression.

I am a researcher from southern Africa, where investment in science is low. What career advice would you give to young scientists who come from countries that are not known for their science research?
When I was a student in Japan the 1960s, I became fascinated by molecular biology. But there was no molecular biology in my country. So I had to go to graduate school abroad. I was fortunate to have the opportunity to study in the United States and I stayed abroad, including in Europe. My antibody work was entirely done in Switzerland. I didn't have much of a relationship with Japan until very recently. But now, after many years, I try to help Japanese science.

For a while, you may have to be trained outside of Africa. Since you are young, if you really want to do science I think there will be a way for you to go abroad and receive training. Then you could go back to southern Africa and try to help science in your country; if you get good training your knowledge will become useful not only to you, but eventually also to the community in which you grew up.

What are your interests outside of science?
I do not have anything that interests me as much as science. But, later in my life I was introduced to music by my wife and our children. I go to concerts with them and I enjoy it. My daughter plays the violin quite well. My younger son played cello and piano. I am not sure what I will focus on if I stop working in the lab. Once in a while I wonder what I am going to do if and when I retire. Is there such a thing as retirement for me? I can't imagine it. ∎

This interview has been edited for length and clarity.

Keikantse Matlhagela is an HIV researcher and lecturer at the University of Botswana's Faculty of Medicine. Previously, she was a Fogarty Research Fellow at the Harvard School of Public Health.

Q&A Elizabeth Blackburn

End-game winner

Elizabeth Blackburn shared the 2009 Nobel Prize in Physiology or Medicine with Carol Greider and Jack Szostak for their work on telomeres — the protective caps at the end of chromosomes — and for identifying the enzyme telomerase, which maintains telomere length. Now at the University of California, San Francisco, she offers Elena Tucker an insight into her life inside and outside academia.

Why did you choose to study the ends of chromosomes?
The driver for me was wanting to understand how life works, rather than solving a particular problem that afflicts humans. When I was finishing my doctoral work at Fred Sanger's lab in Cambridge, UK, DNA-sequencing methods were embryonic. At the time, in the early 1970s, it was hard to sequence DNA except at the ends of relatively short DNA molecules. That was just what was possible. Researchers had looked at the DNA of viruses such as bacteriophages, but I wanted to know what goes on inside the nuclei of cells with real chromosomes. I heard that Joe Gall at Yale University in New Haven, Connecticut, had discovered very short, linear chromosomes in the cell nucleus of the eukaryotic protozoan *Tetrahymena*, and I thought I might be able to sequence the ends of those with the limited technology available at the time.

How did your research get off the ground?
Once at Yale I started analysing the ends of the short *Tetrahymena* chromosomes. What I found was really odd. I had expected to see something similar to the single-strand overhanging DNA observed at the ends of bacteriophage linear DNA, but telomeres were different. They consisted of tandem repeats of a short sequence, and the numbers

⏎ **NATURE.COM**
For an animation about Blackburn and her work visit: go.nature.com/skbuwi

of these repeats varied between molecules and over time. This changeability was a really strong clue of enzyme involvement in maintaining them. Once I had my own lab at the University of California, Berkeley, I extended the telomere research to yeast, collaborating with Jack Szostak at Harvard Medical School in Boston, Massachusetts. We demonstrated that this changeability is not unique to one kind of organism, and that it might be universal to eukaryotes. So I started searching for a new kind of enzyme. That was when my then-graduate student Carol Greider and I discovered telomerase.

We also found that we could make cells quite miserable if we disrupted their telomeres. The molecular puzzles became more and more intriguing. Moving to the University of California San Francisco with its medical school has led me towards all sorts of systems-wide questions around human health and telomeres. I started collaborations to study how chronic stress can affect physiology, which I would never have dreamed of asking on my own.

Why did you decide to enter, and then exit, the start-up world?
Telomere length is not usually by itself diagnostic, but more a general measure of health. Because of the correlation between long

telomeres and a long and healthy life, over the years people have realized that there may be some benefit in measuring the length of their telomeres. Our lab has been quite good at that ever since telomeres have been on the radar. We realized that one day, just as you might send off a DNA sample for sequencing, you might also send a sample to a company that could measure your telomeres. I set up Telome Health with a couple of others, but when it started going in directions that I thought were not scientifically driven I gave away all my equity to the University of California — as did my co-founder Elissa Epel. All in all, the adventure of starting something and talking with people in the finance world was a positive experience. I learned to admire the skills that business people need, realizing that scientists have a lot to learn from their world.

"We found that we could make cells quite miserable if we disrupted their telomeres."

What are the main unanswered questions involving telomeres?
I would like to know how a telomere really works. We have a parts list, so we know what it does in a static sense. But in live cells, telomeres are extraordinarily dynamic. They are complex little ecosystems that constantly have proteins arriving and leaving every second. I think we could learn a huge amount by studying telomeres in action, rather like researchers do by watching active ribosomes assemble proteins, in addition to knowing their structure. The second unanswered question is how we might be able to modify telomere maintenance to make human bodies more resilient, given the correlation between having long telomeres and living healthily and for a long time. I would love to see humanity achieve a situation in which, although we get old and decrepit and things go wrong, we can improve health in the elderly far more than is achieved today.

What happened during your time on the President's Council on Bioethics, which was established by former US president George W. Bush and disbanded by Barack Obama?
When you are a scientist, you are part of this community so you serve on a lot of committees, and sometimes they are National Advisory Commissions. I knew the Bush council had political implications when I joined. I kept being a nag about getting the science right in reports, and they threw me off! That got a lot of attention because it occurred in the context of other developments that spoke to the issue of the use of evidence in policymaking. There were many things at stake. For example, some in government at that time minimized the evidence for climate change. I do believe that although there might be other considerations besides scientific evidence in policymaking,

politically impartial research should always be the bedrock of policy decisions.

Apart from your Nobel prize, you've received a UNESCO L'Oreal for Women in Science Award and been one of *Time* magazine's 100 most influential people. What achievement are you most proud of?
What I'm truly proud of is having done the science and I'm proud of the people who work with me — I'm proud that we've done these things together. The awards are really just symbols. But in the sense that symbols influence people in subtle ways, they matter. If I'm photographed receiving an award, the picture makes the point that there's a woman winning a science prize — that aspect of an award is important to me.

How has your experience of motherhood influenced your career and vice versa?
For years I honestly didn't think much about having children. People would often say "you'd better have a baby soon if you're turning 30", but I didn't listen. I became pregnant in my late thirties when my career was very much underway — I found out I was pregnant in the same week that I was promoted to full professor at Berkeley! I feel very lucky, but I don't think my path is necessarily a recipe for happiness that others should follow. Children can happen at any time, and the challenges will be different depending on where you are in your career.

When you were a child, did your interest in science make you feel different from your peers?
I always felt like a fish out of water. I had good friends growing up and we would do stuff that was normal in Tasmania, Australia, like swim after school and then eat disgustingly delicious meat pies slathered with tomato sauce. But I wouldn't yap to my friends about science. I would talk about Beatles songs. I wasn't afraid of being thought weird, though talking science seemed to come across as pretentious. No one was overtly nasty, but I did sometimes feel pushed away, and that was hurtful. Self-preservation is a useful life lesson. I suppose I always knew that I was different. I was thrilled with the idea that I could go and do a PhD, and do it outside Australia to really expand my horizons. ∎

This interview has been edited for length and clarity.

Elena Tucker is a Peter Doherty Fellow at the Murdoch Childrens Research Institute, Australia, where she studies the genetic basis of rare human disorders.

Q&A Richard Roberts
Microbe cheerleader

Richard Roberts shared the 1993 Nobel Prize in Physiology or Medicine with Phillip Sharp for their discoveries of split genes, which contain parts that encode protein, called exons, and gaps between them, called introns. Now chief scientific officer at New England Biolabs based in Ipswich, Massachusetts, Roberts talks to Gijsbert Werner about microbes, genetically modified food and the problem with Nobel prizes.

In your Lindau lecture this year you talked about genetically modified organisms (GMOs). Are people right to worry about them?

Frankly, they are not. We have been genetically modifying everything we eat for more than 5,000 years. We have been improving plants by 'natural' breeding since the origin of agriculture. When we breed plants, we make hybrids — and typically move hundreds of genes from one plant to another. You don't know what those genes are. You don't know where they go. And you don't know how these genes are influenced by moving them. Genetic engineering is just a better way of doing what we have been doing for the past 5,000 years. The argument that inserting bacterial genes into plants is a break with the past is invalid because, to pick an example, there is very good evidence that the sweet potato genome contains bacterial genes. It doesn't make sense to think that new methods of altering plant genomes will be inherently dangerous. Genes are genes; it is what they do that matters. We need to test whether the products are safe, not worry about the process of creating them. This argument extends to the potential ecosystem effects of GMOs. I do worry about ecosystems, but there is no special risk to them from plants created using these new methods.

One of your main interests is microbes — indeed you gave a lecture about why we should love them at Lindau last year. Why did you feel this was necessary?

The vast majority of the microbes that live with us are good. But bacteria have a bad reputation because science has focused on the ones that cause disease. Biologists are finally starting to realize that by manipulating and controlling microorganisms, we can probably do more for human health than by any other means. The nice thing about this kind of medicine is that it would be cheap. We should explore all sorts of ways to make bacteria more beneficial, including genetic engineering. If you can cure disease by manipulating the microbiome, that is going to save a lot of money and will probably also teach us how to live better. I love bacteria.

Has biotechnology focused too much on the health of the human host without considering its microbial colonizers?

I absolutely think we have gone overboard in studying humans as humans. We need to study good bacteria in the context of their human ecosystems. Until recently, microbiologists did almost no work on good bacteria, which means that these organisms are under-appreciated even though they are an incredibly important part of us. That is a big mistake. The average human contains two to five pounds of bacteria! They provide protection against pathogens and prime our immune systems. If I were to kill all the bacteria that live in or on you, you would probably die. It is as simple as that. We know this because bacteria-free individuals of other species die young.

Why are you so passionate in your support of GM food?

I feel that scientists need to provide more legitimacy to GMOs. A lot of people cannot grasp the nuances of the relevant science, but respect and listen when prominent scientists — particularly Nobel laureates — speak up. I want to make sure the general public receives the benefits of GM food, but also understands its limitations. The fabrications that the anti-GMO people have used to scare the population worry me very much. I would really like to convince green parties of the benefits of GMO. In general, I support green parties. I think they just made a mistake in opposing GM foods — and they did it not because they were against genetic modification per se, but because they were afraid that multinationals were going to take over the food supply.

New techniques are making gene technology available to much smaller organizations than ever before. If what the anti-GMO lobby really cares about is multinationals taking over, might these techniques increase acceptance of GMOs?

The way to think about this is to consider evolution as a very slow process. Plants might eventually adapt to global warming, but if they don't adapt fast enough we won't have enough to eat. Genetic modification is a fast way of doing things. If we do not interfere and 'help' evolution where we can, an awful lot of people are going to die unnecessarily, particularly in the developing world. There are opportunities to really get something done here, and there are strong moral arguments. And there is no reason

why small companies or non-governmental organizations cannot make a big impact and significantly help the developing world.

As well as the new GMO initiative, you also signed the Mainau Declaration on climate change and campaigned in China for the release of Nobel peace laureate Liu Xiaobo. Do you consider it a responsibility to use your Nobel laureate status for the public good?

A Nobel prize is something rather special. Almost all of the laureates here in Lindau were awarded a Nobel prize because we were lucky. It is not that we are super smart or better than anybody else, but because we made a serendipitous discovery along the way. For whatever reason, when you win a Nobel prize people listen to you who never listened before. That means two things. The first is that you should use the opportunity to do good in the world, if you can. The second is that you should also be careful about what you say because you might not always be right. There are plenty of issues in which Nobel laureates could have been helpful, but they were rarely politically organized in the past. We tried to get Aung San Suu Kyi released from house arrest in Myanmar. Even though that was not successful, it showed that we laureates can come together — 225 of us signed letters that were sent to the Chinese and Burmese governments.

> *"Genetic engineering is just a better way of doing what we have been doing for 5,000 years."*

What is the future of the Nobel prizes in the era of big collaborative science, in the light of projects such as ENCODE, the Encyclopedia of DNA Elements?

Many of the major steps forward in biology have been made by individuals or small groups of individuals. Our knowledge of biology is so limited, we are still at the starting point of understanding how organisms work and there are still terrific roles for individuals. But, in general, I am not sure science prizes are a particularly good thing. They are wonderful for the people who win them, and can be terrible for those who don't. I think they end up causing rather a lot of heartbreak. ∎

This interview has been edited for length and clarity.

Gijsbert Werner is a PhD student at Vrije University Amsterdam, the Netherlands, where he studies the evolution of plant-microbe mutualisms.

Q&A Bruce Beutler

Chance encounters

Bruce Beutler is director of the Center for the Genetics of Host Defense at UT Southwestern in Dallas, Texas. He shared one half of the 2011 Nobel Prize in Physiology or Medicine with Jules Hoffmann for their work on the activation of innate immunity; the other half of the prize was awarded to Ralph Steinman. Here, Beutler talks to Christoph Thaiss about biological puzzles and intuition.

The discoveries that have resulted from your work are often referred to as the second revolution in immunology — the elucidation of how innate immunity operates — with the first revolution being adaptive immunity. Will there be a third revolution?

I hope there will be third, fourth and fifth revolutions. People always seem to overestimate what they already know, and we certainly know very little about how the immune system functions. If we think of the immune system as a machine, then we are far from even knowing all of its parts. We cannot predict the outcome of an immune response. We cannot say with confidence who will and who will not get an

autoimmune disease. And we do not know who will and who will not respond to a vaccine. Much more remains to be discovered.

How have the main challenges in medicine changed during your career?

It is interesting that the most challenging diseases — cancer, diabetes and Alzheimer's to name a few — have not really changed since the late 1970s, when I was in medical school, with the exception of emerging infectious diseases. Some conditions are easier to address now, but the major issues remain. Autoimmunity is probably the next frontier. The majority of cases of autoimmune disease result from a complex genetic problem that has environmental influences. It is a colossal task for the immune system to maintain tolerance to self and yet be ready to react to everything in the world around us. We have some ideas about how that works, and we have developed concepts like 'central tolerance' and 'peripheral tolerance' — the two stages by which the immune system learns how to avoid attacking its own body. But when it comes to the mechanisms behind these concepts, we still know very little.

"Today, it is no longer a problem to find mutations that cause phenotypes, which used to be a bottleneck."

What made you decide to use random mutagenesis screens — chance, basically — in your research?

Chance leads to discoveries, and mutagenesis is a way to enhance one's chances of finding a surprise. Often it is the exceptional observations that lead to advances; once you understand exceptions, you understand the whole picture. We used chemically induced random mutagenesis of the mouse genome to identify genes and their functions. A mutation can create an alternative form of a phenomenon — a phenotype or trait — and we can learn a lot by seeing this alternative state. Once I saw a mouse with no eyelids. It simply had a membrane over the eyes. I found it fascinating that there is a single gene required for eyelids to develop. Similarly, one of the most interesting phenotypes I have ever seen is found in a mouse we called 'Possum'. When Possum mice are scruffed at the nape of the neck, they suddenly freeze and go into a sort of trance, for want of a better description. For a few minutes they do not seem to be conscious, but we know from electroencephalograms that they are conscious. The mutation has been identified as a defect in a single type of voltage-gated sodium channel — such a simple cause for such a complicated behavioural phenotype! And remarkably, the channel is not even located within the central nervous system. I find that really puzzling. Mutations get you thinking about how biological processes work.

Another thing I love about mutagenesis is that it is hypothesis-free. I think we can still do good science without having a prediction. If you take hypotheses out of the equation, you also take away the biases that arise because we tend to like our own ideas. If you start with a hypothesis and you find that you were correct, then you cannot really claim to have been surprised. On the other hand, if you start with a phenotype and find the gene that was damaged to create the phenotype, then you can be very much surprised by your discovery.

In that case, how does one develop an intuition for an interesting scientific question?

There is no strict algorithm to follow that leads to interesting discoveries. In my experience, scientists are guided mainly by instinct. In our case, instinct guides the design of screens. In prioritizing phenotypes for study, it helps to ask questions such as: 'Is what we observe unlike anything that has ever been seen before?' and 'Does it have implications for some important aspect of existing theory?'

I get excited by phenotypes that mimic human disease. Today, it is no longer a problem to find the mutations that cause these phenotypes, which certainly used to be the bottleneck in the whole process. Now it takes us about one hour from first seeing a phenotype to finding the causative mutation, and in my lab we usually solve about two phenotypes per day. The difficult part is to understand the mechanism, and there we have to prioritize our experiments so that we learn as much as we can with the resources available.

If there were no technical limitations, what would be your ideal experiment?

I find the speed with which we can already sequence all of an organism's protein-coding genes just magical. The team in my lab is now sequencing about 80 whole exomes — the protein-coding parts of the genome — every two weeks, and I am not sure we need to improve on that much in the future. I feel we have a surfeit of ability, so I am not crying out for new technologies in my own area. But looking more broadly, I think the great technological challenge in medicine in the long term might be in pharmaceutical development. One can envisage a time when we know the three-dimensional structure of all proteins, and that might allow us to compute the structure of drugs that would block certain biological processes without having any side effects. It is an enormous hurdle, but the day may come when computation supplants much of the screening we do presently.

If the scientific system were to be rebuilt from scratch, what do you think it should look like?

It might actually look similar to what we have today. Funding has never been a pure meritocracy, but I do not see a fairer way of doing it, practically speaking. In the area of publishing, I have flirted with the idea that someday there might be no need for peer-reviewed publication. Instead, everyone could publish their best work on their website. Over time, people would learn who the reliable sources were and apply alternative ways of ranking performance. There is a lot of objection to this approach for good reasons. Some are horrified at the prospect that shoddy work would flourish in the absence of peer review. But the model I have in mind would be somewhat similar to the way that artists are evaluated. Who peer-reviewed Bach and his complicated fugues? Imagine what we might have lost if they had been rejected.

Institutional organization is also really important. We need to maintain a mixture of people who work most effectively in small groups, as well as people who are at their best when part of large, organized efforts. Well-coordinated efforts amount to more than the sum of their parts.

If you were not an immunologist, what would you be?

I have always found enormous aesthetic enjoyment in nature. I am an amateur photographer — I take photographs of birds and I like to hike. If I were not a scientist, I might be a naturalist.

But if I were to pick some other field in science, I started out as a neurology resident, and disorders of higher cortical functions in humans still interest me. In the early 1980s, when I began seeing patients, the technology available to study neurobehavioural disease was not nearly as advanced as it is today. The opportunities to understand how the brain works are much greater now. Having a Nobel prize does give me the opportunity to broaden my horizons a bit, and I may move back into neuroscience one day.

What characteristics do you look for in students?

They need to have strong verbal abilities, both written and spoken. I find this to be a predictor of good performance in science in the long run. That may sound strange, but it is an observation I have made many times over the years. ∎

This interview has been edited for length and clarity.

Christoph A. Thaiss is a PhD student in Eran Elinav's group at the Weizmann Institute of Science in Israel. He studies interactions between hosts and their microbiomes, and how these influence susceptibility to common diseases in humans.

Q&A François Englert
Boson beginnings

François Englert shared the 2013 Nobel Prize in Physics with Peter Higgs for the theoretical discovery of a mechanism that gives mass to subatomic particles. For this work, he collaborated with Robert Brout, who died in 2011. He looks back on his contribution to science with Thifhelimbilu Daphney Bucher.

When you were just starting out, what did you find interesting about physics?
Actually, I was an engineer. I started as an engineer, but I also theorized — and I found that more interesting: working out the underlying structure of things rather than their practical applications.

What inspired your early theoretical work?
You have to think back to the situation at the time. At the beginning of the 1960s, physicists understood long-range forces very well. The law of gravity — as described by the general theory of relativity — was proposed by Albert Einstein in 1915. Physicists also understood electromagnetism — the theory of all electric and magnetic phenomena, including the properties of electromagnetic waves. These include light, radio waves, X-rays and γ-rays.

Short-range forces were absolutely not understood back then. So this is what Robert Brout and I, and shortly after also Peter Higgs, initiated — a theory of short-range forces.

What links a theory of short-range forces with the idea of giving subatomic particles mass?
Long-range forces are mediated by the exchange of particles that have no mass. A particle that has no mass, like a photon, travels with the velocity of light. So the idea in our theory of short-range forces was to give particles mass because when they have a mass they do not travel at the speed of light, they are slower than that. And the forces mediated by them become short-range.

How did you and Brout develop the concept of the scalar boson?
At the time, we were studying the theory of phase transition. Phase transition involves a change in the physical properties of matter —such as water becoming vapour or ice. We drew very much from the work of physicist Yoichiro Nambu, who worked on phase transitions and quantum-field theory and showed that they have similarities. This inspired us to introduce to field theory some new particles called scalar bosons, which have the property of being able to form condensates, as occurs in phase transitions.

How does that give particles mass?
The condensate of bosons gives rise to a kind of sea that pervades the whole Universe. Particles travel through that medium, and because of the action of the medium, they may slow down. And so they acquire mass. It is the interaction of the condensate that gives mass, not the individual bosons.

Are there any differences between your theory and the one put forward by Higgs?
There is a difference in the method. Our one — which was inspired by Nambu — is more in line with modern physics than the Higgs approach. At the time, however, people found it more difficult. But our method was well chosen; all of the subsequent development of theory in the field has been done with it.

So why is the term scalar boson less well known than 'Higgs boson'?
That is because of an important paper by US theoretical physicist Steven Weinberg, who shared the Nobel prize in 1979. In a *New York Review of Books*' article in 2012 he discusses this. I can read you a few sentences. It says: "As to my responsibility for the name "Higgs boson," because of a mistake in reading the dates on three earlier papers, I thought that the earliest was the one by Higgs, so in my 1967 paper I cited Higgs first, and have done so since then. Other physicists apparently have followed my lead. But … the earliest paper of the three I cited was actually the one by Robert Brout and François Englert … But the name 'Higgs boson' seems to have stuck."

Does the name of the particle bother you?
When we finished our calculations about scalar bosons, we celebrated because we had worked out something that was mathematically and logically consistent — not because we expected recognition at that time. So I did not care about the name. But the problem is that it is not correct. Correct labelling would be the 'BEH boson' and that would be welcome. I do not wish to complain, however — not now that I have won the Nobel prize! ∎

This interview has been edited for length and clarity.

Thifhelimbilu Daphney Bucher is a South African experimental nuclear physicist. Her research involves developing a γ-ray tracking technique for iThemba LABS's large-volume segmented HPGe detector.